Valószínűségszámítás
1. RÉSZ
Valószínűségek és diszkrét valószínűségi változók

jegyzet

Vetier András

2017. január 27.

Tartalomjegyzék

Tisztelt Hallgatók!

Kérem, hogy a jegyzetben talált hibákat a **vetier@math.bme.hu** email címen jelezzék nekem. A levél tárgya legyen: **Hibát találtam**. Egy hatékony módszer a hibák rögzítésére, ha valaki a számítógépen olvassa a jegyzetet:

- nyom egy PRINT SCREEN-t
- behívja a PAINT programot
- nyom egy PASTE (ctrl-V) utasítást
- pirossal bekarikázza a hibát, esetleg valamit oda ír
- elmenti a JPG fájlt, a fájl neve legyen a hiba helyének az oldalszáma, vagy a nevében legyen benne az oldalszám
- a JPG fájlokat csatolt fájlként a fent megadott címre

Természetesen minden más módszerrel küldött hibajelzést is köszönök.

Az "***Extra feladat***", "***Extra tananyag***" cimkékkel megjelölt feladatok és anyagrészek nem részei a kötelező vizsga anyagának. Ezeket a hétfői vizsgákon nem kérdezzük. Viszont ha valaki az *"Extra feladat"*, *"Extra tananyag"* cimkékkel megjelölt feladatokból és anyagrészekből

- összesen 10 oldalnyi mennyiséget szépen, okosan megtanult, és
- a hétfői vizsga után kedden délig a szándékét emailben jelzi, és
- a szerdai szóbeli beszámoló elején deklarálja, hogy mik azok az anyagrészek, melyekből összeáll az ő extra 10 oldala, és
- utána a vizsgáztató által a deklarált anyagból feltett kérdésekre szépen, okosan válaszol, akkor

a hétfői vizsgán már elért a közepes (3) vagy a jó (4) osztályzatát egy jeggyel javíthatja.

A "szépen, okosan" itt azt jelenti, hogy tényleg érti a dolgokat, nem csupán visszamondja az olvasottakat. iMSc hallgatók a jegy javításán túl iMSc extra pontokat szerezhetnek, ha a megtanult anyag mennyisége 15 oldal. A kedd délig küldendő email címzettje: **vetier@math.bme.hu**, tárgya: **Beszámoló Extra tananyagból**.

Elképzelhető, hogy jogos lenne további részeket is megjelölni az *"Extra feladat"*, *"Extra tananyag"* cimkékkel. Az ilyen jellegű, komolyan átgondolt véleményeiket a **vetier@math.bme.hu** címre írják meg. A levél tárgya legyen: **Javaslat Extra tananyagra**.

Eredményes, örömteli, jó tanulást kívánok! Üdvözlettel,

2016. december 15.

Vetier András

1. Esemény, valószínűség

1.1. Kimenetelek

Véletlen jelenség: Adott körülmények között valami történik (Például két szabályos dobódobókockát szabályosan feldobunk).

Kísérlet: A jelenség önmagától vagy az én szándékom miatt lezajlik. (Korrekt módon gurítom a két dobókockát).

Megfigyelés: Megfigyeljük azt, ami érdekel minket. (Megfigyeljük a dobott számok összegét, vagy szorzatát, hányadosát, de megfigyelhetném azt is, hogy a dobókocka mennyi ideig gurul, hogy hol áll meg, stb.).

Kimenetelek (lehetséges kimenetelek, elemi események): A megfigyelésünk lehetséges eredményei. (Két dobókockával dobva az összeg lehetséges eredményei: 2, 3, ... , 12).

Eseménytér: Az összes lehetséges kimenetelek halmaza. (A példánkban az eseménytér a 11 elemű $\{2, 3, ., 12\}$ halmaz).

1.2. Esemény

Esemény: Állítás a jelenséggel kapcsolatban, ami minden kísérletnél IGAZ (bekövetkezik) vagy HAMIS (nem következik be).

Kísérletsorozat: Több (egymásra semmi hatást kifejtő) kísérletet hajtunk végre.

1.3. Valószínűség

Gyakoriság: Ahányszor bekövetkezik az esemény.

Relatív gyakoriság: Gyakoriság osztva az összes kísérletek számával.

Valószínűség: Egy adott eseményt vizsgálva azt tapasztaljuk, hogy sok kísérlet esetén a relatív gyakoriság közel van egy olyan értékhez, ami nem függ a véletlentől, hanem az eseményre jellemző. Ezt az értéket nevezzük az esemény valószínűségének. A relatív gyakoriság kevés kísérlet esetén erősen eltérhet a valószínűségtől, de a kísérletszám növekedtével egyre közelebb kerül hozzá.

A valószínűségszámítás megtanít arra, hogy a valószínűségeket kísérletek elvégzése nélkül, elméleti módszerekkel meghatározzuk.

1.4. Műveletek eseményekkel

Események – halmazok (Venn diagram): Az eseményeket az eseménytér (mint "alaphalmaz") részhalmazaival reprezentáljuk. Mindenegyes eseményt a szóbanforgó eseményre nézve kedvező kimenetelek által alkotott részhalmaz reprezentál .

Biztos esemény: Mindig bekövetkezik. Az alaphalmazzal reprezentáljuk, amit S -sel jelölünk. Más jelölések: U, I, Ω.

Lehetetlen esemény: Sohase következik be. Az üres halmazzal reprezentáljuk. A $\emptyset$ jellel jelöljük.

Ellentett esemény ("nem", komplementer): Pontosan akkor következik be, amikor az eredeti esemény nem. Felülvonással jelöljük, pl.: $\overline{A}$.

Események "és" kapcsolata (metszet, közös rész, szorzat): A szóbanforgó események mindegyike bekövetkezik. Jelölések:

Két esemény metszete: $A \cap B$
Véges sok esemény metszete: $A_1 \cap A_2 \ldots \cap A_n$
Végtelen sok esemény metszete: $A_1 \cap A_2 \cap \ldots$

Kizáró események: A szóbanforgó események közül legfeljebb egy következhet be. Bármely kettő kizárja egymást, egyidejű bekövetkezésük lehetetlen.

Két eseményre: $A \cap B = \emptyset$.
Több (véges vagy végtelen sok) eseményre: ha $i \neq j$, akkor $A_i \cap A_j = \emptyset$.

Események "vagy" kapcsolata (únió, egyesítés, összeg): A szóbanforgó események közül legalább egy bekövetkezik. Jelölések:

Két esemény úniója: $A \cup B$
Véges sok esemény úniója: $A_1 \cup A_2 \ldots \cup A_n$
Végtelen sok esemény úniója: $A_1 \cup A_2 \cup \ldots$

Kizáró események "vagy" kapcsolata (úniója, egyesítése, összege): A szóbanforgó események közül pontosan egy bekövetkezik. A jelölések általában nem különböznek a korábbi jelölésektől:

Két kizáró esemény úniója: $A \cup B$
Véges sok kizáró esemény úniója: $A_1 \cup A_2 \ldots \cup A_n$
Végtelen sok kizáró esemény úniója: $A_1 \cup A_2 \cup \ldots$

Azonban kizáró eseménykre bizonyos szabályok egyszerűbbek az általános szabályoknál, vannak olyan könyvek, ahol egy *-gal hívják fel a figyelmet arra a tényre, hogy műveletben szereplő események kizáróak:

Két kizáró esemény úniója: $A \cup^* B$
Véges sok kizáró esemény úniója: $A_1 \cup^* A_2 \ldots \cup^* A_n$
Végtelen sok kizáró esemény úniója: $A_1 \cup^* A_2 \cup^* \ldots$

"Maga után vonja": Azt mondjuk, hogy egy B esemény maga után vonja az A eseményt, ha teljesül, hogy amikor a B esemény bekövetkezik, akkor szükségképpen az A esemény is bekövetkezik, azaz a B halmaz része az A halmaznak. Jelölés:

$B \subset A$ vagy $A \supset B$.

1.5. A valószínűség alapvető tulajdonságai

Ebben az alfejeztben valószínűség alapvető tulajdonságait soroljuk fel, melyek – a relatív gyakoriság ugyanilyen tulajdonságai alalpján – kézenfekvőek. Fontos, hogy az olvasó ne csak memorizálja ezeket a tulajdonságokat, hanem értse, lássa, hogy *a valószínűségnek ezek a tulajdonságai a relatív gyakoriság (triviális!) hasonló tulajdonságai miatt igazak*. Néhány bonyolultabb szabályt egy későbbi fejezetben sorolunk fel. Megjegyezzük, hogy az 1., a 2. és a 5. tulajdonságból logikai úton le lehet vezetni a valószínűség összes többi tuldajdonságát, ezért a valószínűségszámítás axiómatikus felépítésekor ezek szolgálnak axiómákként. Ebben a jegyzetben nem célunk az axiómatikus tárgyalás. A technikásabb bizonyításokat is leegyszerűsítve, a szemléletre támaszkodva mutatjuk majd be. Fő célunk az elmélet és a valóság kapcsolatátának világos tálalása.

1. **Minden esemény valószínűsége** 0 **és** 1 **közé esik:**

$$0 \leq \mathrm{P}(A) \leq 1$$

2. **A biztos esemény valószínűsége** 1 **:**

$$\mathrm{P}(S) = 1$$

3. **A lehetetlen esemény valószínűsége** 0 **:**

$$P(\emptyset) = 0$$

4. **Komplementer szabály**:
Minden A eseményre igaz, hogy

$$\mathrm{P}(A) + \mathrm{P}(\overline{A}) = 1 \qquad \mathrm{P}(A) = 1 - \mathrm{P}(\overline{A}) \qquad \mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$$

5. **Összegzési szabály kizáró eseményekre:**
Ha $A_1, A_2, \ldots, A_n$ véges sok kizáró esemény, és $A = A_1 \cup A_2 \cup \ldots \cup A_n$, akkor

$$\mathrm{P}(A) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \ldots + \mathrm{P}(A_n)$$

Ha $A_1, A_2, A_3, A_4, \ldots$ végtelen sok kizáró esemény, és $A = A_1 \cup A_2 \cup A_3 \cup A_4 \cup \ldots$, akkor

$$\mathrm{P}(A) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \mathrm{P}(A_4) + \ldots$$

6. **Általános összegzési szabály** (még csak) **két eseményre**:
Ha A, B tetszőleges események, akkor

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$

7. **Általános kivonási szabály**:
Ha a A, B tetszőleges események, akkor

$$\mathrm{P}(A \backslash B) = \mathrm{P}(A) - \mathrm{P}(A \cap B)$$

8. **Speciális kivonási szabály**:
   Ha a B esemény maga után vonja az A eseményt, vagyis $B \subseteq A$, akkor
$$\mathrm{P}(A \backslash B) = \mathrm{P}(A) - \mathrm{P}(B)$$

Feladat: Páros vagy páratlan? Egy érmét dobálunk az első fejig. Megfigyeljük az ehhez szükséges dobások számát. Mi a valószínűsége annak, hogy az ehhez szükséges dobások száma páros szám?

Megoldás: A lehetséges kimenetelek: $1, 2, 3, 4, 5, 6, \ldots$, soha, ahol a "soha" akkor következik be, ha mindig csak írást dobunk, vagyis soha se dobunk fejet. Az alábbi valószínűségek triviálisak:

$$\mathrm{P}(\text{az első fej a 2. dobásra adódik}) = \tfrac{1}{2^2} = \tfrac{1}{4}$$
$$\mathrm{P}(\text{az első fej a 4. dobásra adódik}) = \tfrac{1}{2^4} = \tfrac{1}{16}$$
$$\mathrm{P}(\text{az első fej a 6. dobásra adódik}) = \tfrac{1}{2^6} = \tfrac{1}{64}$$
$$\vdots$$

Ezekből a feladat kérdésére a válasz összegzéssel adódik:

$$\begin{aligned}
&\mathrm{P}(\text{az első fej eléréséhez páros sok dobás kell}) = \\
&= \mathrm{P}(\text{az első fej a 2. dobásra adódik}) + \\
&+ \mathrm{P}(\text{az első fej a 4. dobásra adódik}) + \\
&+ \mathrm{P}(\text{az első fej a 6. dobásra adódik}) + \\
&\vdots \\
&= \tfrac{1}{4} + \tfrac{1}{16} + \tfrac{1}{64} + \ldots \;=\; \tfrac{\frac{1}{4}}{1-\frac{1}{4}} \;=\; \tfrac{\left(\frac{1}{4}\right)}{\left(\frac{3}{4}\right)} \;=\; \tfrac{1}{3}
\end{aligned}$$

Az utolsó sorban, a végtelen sor összegzésénél, felhasználtuk a jól ismert tényt, hogy egy végtelen mértani sor összegére igaz:
$$\text{összeg} = \frac{\text{első tag}}{1 - \text{kvóciens}}$$

1.6. Klasszikus problémák

Gyakran megesik, hogy a megfigyelésünknek véges sok kimenetele van, melyek (valamilyen szimetria) miatt érezhetően egyforma valószínűségűek. Ilyenkor minden kimenetel valószínűsége a lehetséges kimenetelek számának a reciproka, és egy esemény valószínűsége egyenlő az eseményre nézve kedvező kimenetelek száma osztva az összes események számával:

$$\mathrm{P}(A) = \frac{\textit{az eseményre nézve kedvező kimenetelek száma}}{\textit{az összes kimenetelek száma}}$$

1.7. Számlálási alapszabályok

Ezek a szabályok teljesen nyilvánvalóak, mindenki ismeri őket. Mégis felsoroljuk őket, hogy amikor kell, hivatkozhassunk rájuk.

Összegzés: Ha egy halmaz egymást kizáró részhalmazokra bomlik (a halmazt particionáljuk, a halmaz partíciókra bomlik), akkor *a halmaz elemszáma egyenlő a részhalmazok elemszámainak összegével.*

Kivonás: Ha egy halmaznak elhagyjuk egy részhalmazát, akkor *a megmaradó halmaz elemeinek száma egyenlő az eredeti halmaz elemszáma mínusz az elhagyott részhalmaz elemszáma.*

Szorzás: Ha két halmazt tekintünk, és az egyiket "első"-nek nevezzük, a másikat "második"-nak, és rendezett párokat képezünk az elemeikből úgy, hogy az első halmazból vesszük a párok első elemeit, a másodikból a párok második elemeit, akkor *a párok darabszáma egyenlő a két halmaz elemszámának a szorzatával.*

Több tényezős szorzás: Ha n halmazt tekintünk, és az egyiket "első"-nek nevezzük, a másikat "második"-nak, és így tovább kapjuk az $(n-1)$ -ik, n -ik halmazokat, és rendezett n -eseket képezünk az elemeikből úgy, hogy az első halmazból vesszük az első elemet, a másodikból a másodikat, és így tovább az n -ik halmazból vesszük az n -ik elemet, akkor *a rendezett n -esek darabszáma egyenlő a halmazok elemszámainak a szorzatával.*

Több tényezős szorzás fa-gráfokkal: Képzeljünk el egy fa-gráfot, mely "felfelé nő", és
gyökeréből k_1 él indul ki (ezek az elsőrendű élek),
az elsőrendű élek mindegyikének a felső végpontjából k_2 él indul ki (ezek a másodrendű élek),
a másodrendű élek mindegyikének a felső végpontjából k_3 él indul ki (ezek a harmadrendű élek),
és így tovább,
az $(n-1)$ -ed rendű élek mindegyikének a felső végpontjából k_n él indul ki.
E fa-gráf tetején a végpontok száma: $k_1 \cdot k_2 \cdot k_3 \cdot \ldots \cdot k_n$.

Hatványozás: Ha egy halmazt n példányban tekintünk, és az egyiket "első"-nek nevezzük, a másikat "második"-nak, és így tovább kapjuk az $(n-1)$ -ik n -ik példányát a halmaznak, és rendezett n -eseket képezünk az elemeikből úgy, hogy az első halmazból vesszük a az első elemet, a másodikból a másodikat, és így tovább az n -ik halmazból vesszük az n -ik elemet, akkor *a rendezett n -esek darabszáma egyenlő a halmaz elemszámának n -ik hatványával.*

Osztás: Ha egy halmazt úgy particionálunk (bontunk diszjunkt részhalmazokra), hogy minden partició (részhalmaz) ugyanannyi elemből áll, akkor *a particiók (részhalmazok) darabszáma egyenlő a halmaz elemszáma osztva a partíciók (részhalmazok) közös elemszámával.*

1.8. Kombinatorikus alapképletek

Az alábbi táblázatba folgalt képleteket ismertnek feltételezzük. Egy-egy példával világítunk rá jelentésükre.

	ismétlés nélküli	ismétléses
permutáció	$n!$ n futó beérkezésének sorrendje	$\frac{n!}{k_1!k_2!\dots k_r!}$ n golyót ennyiféleképpen állíthatunk sorba, ha $k_1, k_2, \dots k_r$ db külön-külön egyszínű
variáció	$\frac{n!}{(n-k)!}$ n futó beérkezésének sorrendje ha csak az első k helyet tekintjük	l^k l darab betűből készíthető k hosszú szavak száma (egy-egy betű többször is felhasználható)
kombináció	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ n golyóból kiválasztunk k darabot (nem számít a kiválasztás sorrendje)	$\binom{k+l-1}{l}$ k féle sütiből (sok van belőlük) hazaviszünk l -et, ennyiféleképpen tehetjük meg *(Extra tananyag)*

Megjegyzés: Az ismétléses kombináció képletét meg lehet említeni, de nem kell foglalkozni vele.

1. Példa: Lottó öt találat. Az ötös lottón 90 szám közül húznak ki 5-öt. A nyerés szempontjából a sorrend nem számít, ezért az összes lehetséges kombinációk száma

$$\binom{90}{5} = 43,949,268 \approx 44 \text{ millió}$$

A biztos teli találat eléréséhez ennyi szelvényt kellene kitöltenünk. Megemlítjük, hogy ha 44 millió lottószelvényt egymásra raknánk, akkor ez a torony a Föld legmasabb csúcsáig, a Mount Everest tetejéig érne fel. Ha egyetlen szelvénnyel játszom, akkor az öt találatom valószínűsége

$$\frac{1}{\binom{90}{5}} = \frac{1}{43,949,268} \approx 0.000000023$$

hiszen a 43,949,268 egyformán valószínű kombináció között az az egyetlen a kedvező, ahogyan én töltöm ki a szelvényt.

2. Példa: Lottó találatok. Annak az eseményeknek a valósznűsége, hogy egy szelvénnyel játszva az ötös lottón, a találataim száma k, az alábbi törttel adható meg:

$$\frac{\binom{5}{k}\binom{85}{5-k}}{\binom{90}{5}}$$

ahol $k = 0, 1, \dots, 5$. A valószínűségek numerikus értéke:

$$\text{P(5 találotos lesz a szelvényem)} = \frac{\binom{5}{5}\binom{85}{0}}{\binom{90}{5}} \approx 0.000000023$$

$$\text{P(4 találotos lesz a szelvényem)} = \frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}} \approx 0.000009670$$

$$\text{P(3 találotos lesz a szelvényem)} = \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} \approx 0.000812300$$

$$P(\text{ 2 találotos lesz a szelvényem }) = \frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}} \approx 0.022473639$$

$$P(\text{ 1 találotos lesz a szelvényem }) = \frac{\binom{5}{1}\binom{85}{4}}{\binom{90}{5}} \approx 0.230354804$$

$$P(\text{ 0 találotos lesz a szelvényem }) = \frac{\binom{5}{0}\binom{85}{5}}{\binom{90}{5}} \approx 0.746349564$$

3. Példa: Hány piros? A lottó probléma általánosítása: tegyük fel, hogy egy dobozban N darab golyó van. Közülük K darab piros, $N-K$ darab fehér. Kiveszünk n darab golyót visszatevés nélkül. Mi a valószínége annak, hogy a kivett golyók között pontosan k darab piros lesz? Válasz:

$$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

4. Példa: Hány piros, hány kék? Az előző probléma általánosítása: tegyük fel, hogy egy dobozban N darab golyó van. Közülük K_1 darab piros, K_2 darab kék, $N-K_1-K_2$ darab fehér. Kiveszünk n darab golyót visszatevés nélkül. Mi a valószínége annak, hogy a kivett golyók között pontosan k_1 darab piros és pontosan k_2 darab kék lesz? Válasz:

$$\frac{\binom{K_1}{k_1}\binom{K_2}{k_2}\binom{N-K_1-K_2}{n-k_1-k_2}}{\binom{N}{n}}$$

5. Példa: Nyolcszor húzunk. Az előző példában feltett általános kérdést most egy speciális esetben alaposabban megvizsgáljuk. Tegyük fel, hogy 45 darab golyó van egy dobozban. Közülük 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 darab golyót visszatevés nélkül. Mi a valószínége annak, hogy a kivett golyók között pontosan x darab piros és pontosan y darab kék lesz? Válasz:

$$\frac{\binom{10}{x}\binom{15}{y}\binom{20}{8-x-y}}{\binom{45}{8}}$$

minden olyan x -re és y -ra, melyek eleget tesznek a következő egyenlőtlenségeknek:

$$0 \leq x \leq 8$$

$$0 \leq y \leq 8$$

$$0 \leq 8-x-y \leq 8$$

A valószínűségek numerikus értékei segítségével többször fogunk majd a későbbiekben dolgozni, ezért Excellel kiszámoltuk, és táblázatba rendezve itt megadjuk őket:

y										
8	0.000									
7	0.001	0.000								
6	0.004	0.005	0.001							
5	0.016	0.026	0.013	0.002						
4	0.0317	0.072	0.054	0.015	0.001					
3	0.033	0.102	0.108	0.048	0.009	0.001				
2	0.019	0.0756	0.106	0.067	0.019	0.002	0.000			
1	0.005	0.027	0.049	0.040	0.017	0.003	0.000	0.000		
0	0.001	0.004	0.008	0.009	0.005	0.001	0.000	0.000	0.000	
	0	1	2	3	4	5	6	7	8	x

1.9. RANDBETWEEN utasítás

Az Excelben az egész értékeket felvevő

`RANDBETWEEN(A;B)`,   magyarul `VÉLKÖZÖTT(A;B)`

utasítás olyan véletlen számot állít elő, mely egyenletes eloszlást követ az $\{A, A+1, \ldots, B-1, B\}$ halmazon.

1. Példa: Száz cédula. A

`RANDBETWEEN(1;100)`

utasítással azt szimulálhatjuk, mintha az $\{1, 2, \ldots, 99, 100\}$ számokat egy-egy cédulára írnánk, a cédulákat egy dobozba tennénk, és a dobozból kihúznánk egy cédulát, és megnéznénk a rajta lévő számot. Annak a valószínűsége, hogy `RANDBETWEEN(1;100)` értéke

pontosan 55, egyenlő $1/100$ -dal

kisebb vagy egyenlő, mint 55, egyenlő $55/100 = 0.55$ -dal

nagyobb 50 -nél, de kisebb 60 -nál, egyenlő $9/100 = 0.09$ -dal

2. Példa: Dobókocka. A

`RANDBETWEEN(1;6)`

utasítással azt szimulálhatjuk, mintha egy szabályos dobókockával dobnánk, és tekintenénk a dobott számot. A 6 lehetséges eset mindegyike $\frac{1}{6}$ valószínűségű.

Fontos, hogy Olvasó tisztában legyen azzal, hogy ha a `RANDBETWEEN(1;6)` utasítást többször leírjuk, akkor minden alkalmazás a többitől független eredményt ad. Ha az utasítást ahhoz hasonlóan, ahogy most ideírjuk:

`RANDBETWEEN(1;6)`   `RANDBETWEEN(1;6)`

két külön Excel-cellába is beírjuk, azt szimulálhatjuk, mintha két szabályos dobókockával dobnánk.

3. Példa: Két dobókocka. Két szabályos dobókockával dobunk. A két dobókockát (még akkor is, ha teljesen egyformának tűnnek) meg tudjuk különböztetni, ha az egyiket a bal, a másikat a jobb kezünkből gurítjuk. A két dobókocka ily módon való dobását szimulálhatjuk a

`RANDBETWEEN(1;6)`   `RANDBETWEEN(1;6)`

utasításpárral. Nyilván 36 lehetséges kimenetel kínálkozik. Ezt a 36 esetet – egymás után leírva – fel is sorolhatjuk, de előnyös, ha táblázatba rendezve adjuk meg őket. Íme:

jobb bal	1	2	3	4	5	6
1	1 ; 1	1 ; 2	1 ; 3	1 ; 4	1 ; 5	1 ; 6
2	2 ; 1	2 ; 2	2 ; 3	2 ; 4	2 ; 5	2 ; 6
3	3 ; 1	3 ; 2	3 ; 3	3 ; 4	3 ; 5	3 ; 6
4	4 ; 1	4 ; 2	4 ; 3	4 ; 4	4 ; 5	4 ; 6
5	5 ; 1	5 ; 2	5 ; 3	5 ; 4	5 ; 5	5 ; 6
6	6 ; 1	6 ; 2	6 ; 3	6 ; 4	6 ; 5	6 ; 6

A 36 eset mindegyike $\frac{1}{36}$ valószínűségű:

bal \ jobb	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

4. Példa: Két dobókockával dobott számok összege. Egyes társasjátékokban két dobókockával dobunk, és a játékban a dobott számok összege, azaz – a szimuláció nyelvén mondva – a

```
RANDBETWEEN(1;6)+ RANDBETWEEN(1;6)
```

utasítás értéke számít. A következő táblázatban at összeg értékeit tüntetjük fel:

bal \ jobb	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

A táblázat 36 cellájában

2 -es érték	1 -szer
3 -as érték	2 -szer
4 -es érték	3 -szor
5 -ös érték	4 -szer
6 -os érték	5 -ször
7 -es érték	6 -szor
8 -as érték	5 -ször
9 -es érték	4 -szer
10 -es érték	3 -szor
11 -es érték	2 -szer
12 -es érték	1 -szer

szerepel. Ezért annak a valószínűsége, hogy az összeg értéke

$$
\begin{array}{rll}
2\,, & \text{egyenlő} & 1/36 \text{ -dal} \\
3\,, & \text{egyenlő} & 2/36 \text{ -dal} \\
4\,, & \text{egyenlő} & 3/36 \text{ -dal} \\
5\,, & \text{egyenlő} & 4/36 \text{ -dal} \\
6\,, & \text{egyenlő} & 5/36 \text{ -dal} \\
7\,, & \text{egyenlő} & 6/36 \text{ -dal} \\
8\,, & \text{egyenlő} & 5/36 \text{ -dal} \\
9\,, & \text{egyenlő} & 4/36 \text{ -dal} \\
10\,, & \text{egyenlő} & 3/36 \text{ -dal} \\
11\,, & \text{egyenlő} & 2/36 \text{ -dal} \\
12\,, & \text{egyenlő} & 1/36 \text{ -dal}
\end{array}
$$

egyenlő. Ha az összeg lehetséges értékeit és a kapott valószínűségeket táblázatba rendezzük, az alábbi táblázatot kapjuk:

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

5. Példa: Két dobókockával dobott számok szorzata. Ha valakit a dobott számok szorzata érdekel, akkor az elméleti okoskodáshoz olyan táblázatra van szüksége, mely a szorzatokat tartalmazza:

jobb / bal	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

A táblázat 36 cellájában

1 -es érték	1 -szer
2 -es érték	2 -szer
3 -as érték	2 -szer
4 -es érték	3 -szor
5 -ös érték	2 -szer
6 -os érték	4 -szer
8 -as érték	2 -szer
9 -es érték	1 -szer
10 -es érték	2 -szer
12 -es érték	4 -szer
15 -ös érték	2 -szer
16 -es érték	1 -szer
18 -as érték	2 -szer
20 -as érték	2 -szer
24 -es érték	2 -szer
25 -es érték	1 -szer
30 -es érték	2 -szer
36 -es érték	1 -szer

szerepel. Ezért annak a valószínűsége, hogy a szorzat értéke

1 ,	$1/36$ -dal egyenlő
2 ,	$2/36$ -dal egyenlő
3 ,	$2/36$ -dal egyenlő
4 ,	$3/36$ -dal egyenlő
5 ,	$2/36$ -dal egyenlő
6 ,	$4/36$ -dal egyenlő
8 ,	$2/36$ -dal egyenlő
9 ,	$1/36$ -dal egyenlő
10 ,	$2/36$ -dal egyenlő
12 ,	$4/36$ -dal egyenlő
15 ,	$2/36$ -dal egyenlő
16 ,	$1/36$ -dal egyenlő
18 ,	$2/36$ -dal egyenlő
20 ,	$2/36$ -dal egyenlő
24 ,	$2/36$ -dal egyenlő
25 ,	$1/36$ -dal egyenlő
30 ,	$2/36$ -dal egyenlő
36 ,	$1/36$ -dal egyenlő

egyenlő. Ha a szorzat lehetséges értékeit és a kapott valószínűségeket táblázatba rendezzük, az alábbi táblázatot kapjuk:

1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

6. Példa: Két dobókockával dobott számok hányadosa. Ha valakit a dobott számok hányadosa (mondjuk, a bal kézből gutított dobókockán lévő szám osztva a jobb kézből gurított dobókockán lévő szám) a érdekel, akkor az elméleti okoskodáshoz olyan táblázatra van szüksége, mely a hányadosokat tartalmazza:

jobb / bal	1	2	3	4	5	6
1	1	1/2	1/3	1/4	1/5	1/6
2	2	1	2/3	1/2	2/5	1/3
3	3	3/2	1	3/4	3/5	1/2
4	4	2	4/3	1	4/5	2/3
5	5	5/2	5/3	5/4	1	5/6
6	6	3	2	3/2	6/5	1

A táblázat 36 cellájában az

1/6 érték	1 -szer
1/5 érték	2 -szer
1/4 érték	2 -szer
1/3 érték	3 -szor
2/5 érték	2 -szer
1/2 érték	4 -szer
3/5 érték	2 -szer
2/3 érték	1 -szer
3/4 érték	2 -szer
4/5 érték	1 -szer
5/6 érték	2 -szer
1 érték	1 -szer
6/5 érték	2 -szer
5/4 érték	2 -szer
4/3 érték	2 -szer
5/3 érték	1 -szer
2 érték	1 -szer
5/2 érték	1 -szer
3 érték	1 -szer
4 érték	1 -szer
5 érték	1 -szer
6 érték	1 -szer

szerepel. Ezért annak a valószínűsége, hogy a hányados értéke

1/6 , egyenlő 1/36 -dal
1/5 , egyenlő 2/36 -dal
1/4 , egyenlő 2/36 -dal
1/3 , egyenlő 3/36 -dal
2/5 , egyenlő 2/36 -dal
1/2 , egyenlő 4/36 -dal
3/5 , egyenlő 2/36 -dal
2/3 , egyenlő 1/36 -dal
3/4 , egyenlő 2/36 -dal
4/5 , egyenlő 1/36 -dal
5/6 , egyenlő 2/36 -dal
1 , egyenlő 1/36 -dal
6/5 , egyenlő 2/36 -dal
5/4 , egyenlő 2/36 -dal
4/3 , egyenlő 2/36 -dal
5/3 , egyenlő 1/36 -dal
2 , egyenlő 1/36 -dal
5/2 , egyenlő 1/36 -dal
3 , egyenlő 1/36 -dal
4 , egyenlő 1/36 -dal
5 , egyenlő 1/36 -dal
6 , egyenlő 1/36 -dal

Ha a hányados lehetséges értékeit és a kapott valószínűségeket táblázatba rendezzük, az alábbi táblázatot kapjuk:

1/6	1/5	1/4	1/3	2/5	2/4	3/5	2/3	3/4	4/5	5/6	...
1/36	1/36	1/36	2/36	1/36	3/36	1/36	2/36	1/36	1/36	1/36	...

...	1	6/5	5/4	4/3	3/2	5/3	2	5/2	3	4	5	6
...	6/36	1/36	1/36	1/36	2/36	1/36	3/36	1/36	2/36	1/36	1/36	1/36

1.10. Gyakorló feladatok

I. téma: Lehetséges kimenetelek

Az alábbi véletlen jelenségek megnevezett megfigyelésével kapcsolatban adjuk meg az eseményteret, azaz soroljuk fel a lehetséges kimeneteleket (más néven: elemi eseményeket). Minden esetben állapítsuk meg, hogy hány elemű az eseménytér?

1. (a) Két szabályos érmével dobunk,
   (b) Három szabályos érmével dobunk,

   (c) Négy szabályos érmével dobunk,
   (d) Öt szabályos érmével dobunk,
   (e) Tíz szabályos érmével dobunk,

   és megfigyeljük mindegyik érmén, hogy melyik oldal van felül.

2. (a) Addig dobunk egy szabályos érmével, amíg végre fejet kapunk,
   (b) Addig dobunk egy szabályos érmével, amíg másodszorra fejet kapunk,
   (c) Addig dobunk egy szabályos érmével, amíg harmadszorra fejet kapunk,

   és megfigyeljük, hány dobás kell ehhez.

3. (a) Addig dobunk egy szabályos érmével, amíg végre fejet kapunk,
   (b) Addig dobunk szabályos érmével, amíg másodszorra fejet kapunk,
   (c) Addig dobunk szabályos érmével, amíg harmadszorra fejet kapunk,

   és megfigyeljük, hogy addig hányszor dobtunk írást.

4. (a) Két szabályos (különböző színű) dobókockával dobunk,
   (b) Három szabályos (különböző színű) dobókockával dobunk,
   (c) Négy szabályos dobókockával dobunk,
   (d) Öt szabályos (különböző színű) dobókockával dobunk,

   és megfigyeljük mindegyik kockán, hogy melyik szám van felül.

5. (a) Addig dobunk egy szabályos dobókockával, amíg végre hatost kapunk,
   (b) Addig dobunk egy szabályos dobókockával, amíg másodszorra hatost kapunk,
   (c) Addig dobunk egy szabályos dobókockával, amíg harmadszorra hatost kapunk,

   és megfigyeljük, hány dobás kell ehhez.

II. téma: Kombinatórika gyakorlása

6. A hét törpe minden este más sorrendben szeretne sorba állni, amikor Hófehérke a vacsorát osztja. Hányféleképpen tehetik ezt meg?

7. Hányféle sorrendben rakhatók ki a MATEMATIKA szó betűi?

8. Egy versenyen 5-en indulnak, az újságok az első három helyezett nevét közlik. Hányféle lehet ez a lista? (Közlik a helyezést is.)

9. Egy fagyizóban 5 féle fagylalt kapható: vanília, csoki, málna, pisztácia és citrom. Hányféleképpen vehetünk 2 gombócot, ha számít a gombócok sorrendje is, és lehet egyfajtából többet is venni?

10. Van 6 lányismerősöm, és kettőt el akarok hívni moziba. Hányféleképpen tehetem ezt meg?

11. 3 új tanárt és egy titkárnőt akarnak felvenni egy iskolában. 6 tanár- és 3 titkárnő-jelölt van. Hányféleképpen kerülhetnek ki közülük az iskola új dolgozói?

12. Egy számkombinációs zárat 3 db különböző, 1 és 10 közötti szám begépelésével lehet kinyitni, de tudjuk, hogy a számok növekvő sorrendben vannak. Hány ilyen kombináció van?

III. téma: Klasszikus képlet alkalmazása

13. Feldobunk egy érmét kétszer egymásután. Mi a valószínűsége, hogy dobunk fejet? És hogy pontosan 1 db fejet dobunk?

14. Egy csomag magyar kártyából kiveszünk egy lapot, megnézzük a színét, majd visszatesszük. Megkeverjük a paklit, majd megint választunk egy lapot. Mennyi a valószínűsége annak, hogy a két lap színe különböző?

15. Mi a valószínűsége annak, hogy két darab (szabályos) kocka feldobásakor legalább az egyik 6-os lesz? És annak a valószínűsége, hogy egyik sem lesz 6-os?

16. Mi a valószínűsége annak, hogy egy háromgyermekes családban a gyerekek mind egyneműek, ha a lányok és a fiúk születési valószínűsége egyaránt $\frac{1}{2}$?

17. Legalább hány szabályos pénzdarabot kell feldobni ahhoz, hogy 90%-nál nagyobb legyen az esély arra, hogy legyen köztük fej?

18. Mennyi a valószínűsége, hogy ha egy polcon 7 db könyvet véletlenszerűen sorba rakunk, akkor egy köztük lévő trilógia kötetei egymás mellé kerülnek?

19. Hatszor dobunk egy szabályos dobókockával. Mi a valószínűsége annak, hogy mind a hat szám előjön?

20. A brazil labdarúgó válogatott edzésének megkezdése előtt, az edzésen résztvevő 22 játékost két csoportba osztják. Mi annak a valószínűsége, ha találomra történik a szétosztás a két 11-es csoportba, hogy Ronaldo és Ronaldinho egymás ellen játszik?

21. Mennyi a valószínűsége annak, hogy az ötös lottón (90-ből 5 számot húznak, sorrend nem számít) pontosan két találatunk lesz? És hogy legalább két találatunk lesz?

22. Egy dobozban 6 zöld és 4 sárga golyó van. Kihúzunk (visszatevés nélkül) 4 golyót csukott szemmel, mennyi a valószínűsége, hogy pontosan két zöld golyót húztunk ki?

2. Diszkrét eloszlás

2.1. Valószínűségi változó

Amikor a megfigyelés eredményei, vagyis a lehetséges kimenetelek *számok*, akkor azt mondjuk, hogy **valószínűségi változó**val van dolgunk. A gyakorlati alkalmazásokban felbukkanó valószínűségi változók két fontos csoportba sorolhatók:

1. A valószínűségi változónak véges vagy megszámlálhatóan végtelen sok lehetséges értéke van. Ilyenkor azt mondjuk, hogy a valószínűségi változó **diszkrét**.
2. A valószínűségi változó lehetséges értékei egy véges vagy végtelen intervallumot tesznek ki Ilyenkor azt mondjuk, hogy a valószínűségi változó **folytonos**.

2.2. Eloszlás és súlyfüggvény

Ha egy diszkrét valószínűségi változó minden lehetséges értékének megadjuk a valószínűségét (táblázattal, képlettel, stb.), akkor azt mondjuk, hogy megadjuk a **valószínűségi változó eloszlásá**t.

Ha (mindenféle valószínűségi változóra való tekintet nélkül) egy véges vagy megszámlálhatóan végtelen halmaz x elemeihez nemnegatív $p(x)$ számokat rendelünk úgy, hogy azok összege 1:

$$\sum_x p(x) = 1$$

akkor azt mondjuk, hogy megadunk egy **valószínűségi eloszlás**t (**valószínűség eloszlás**t), más kifejezéssel **normált eloszlás**t. A $p(x)$ függvényt **súlyfüggvény**nek vagy (**valószínűségi függvény**nek) nevezzük.

Egy X valószínűségi változó esetén $p(x)$ megadja annak a valószínűségét, hogy a véletlen X érték x -szel egyenlő:

$$p(x) = \mathrm{P}(X = x)$$

1. Példa: Fiatal házaspárok gyerekei. Tegyük fel, hogy egy szociológiai felmérésben fiatal házaspárok (akik 3 évnél nem régebben házasodtak) gyermekeinek számát vizsgálták. A gyermekek száma 0, 1, 2, 3 lehet. A tapasztalatból tudhatjuk a négy lehetőség mindegyikének százalékos értékét. Tegyük fel, hogy ezek az alábbiak:

gyerekek száma	0	1	2	3
százalék	20	40	30	10

(A négy darab százalék érték összege természetesen 100.)

Ha egy fiatal házaspárt véletlenszerűen választunk, és tekintjük az alábbi valószínűségi változót:

$$X = \text{ gyerekek száma}$$

akkor a négy lehetséges érték mindegyikének a valószínűségét (a megfelelő százalék érték alapján) beírva a táblázatba megkapjuk az X valószínűségi változó eloszlását:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

(A négy darab valószínűség érték összege természetesen 1.)

2. Példa: Két dobókockával dobott számok összege – ismét. Ha két dobókockával dobva a dobott számok összegét tekintjük, akkor az így kapott valószínűségi változó eloszlása:

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

3. Példa: Két dobókockával dobott számok szorzata – ismét. Ha két dobókockával dobva a dobott számok szorzatát tekintjük, akkor az így kapott valószínűségi változó eloszlása:

x	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2.3. Eloszlásfüggvény

Egy adott normált eloszlás esetén egy x lehetséges értékkel kapcsolatban a $(-\infty, x]$ intervallum valószínűségét $F(x)$ -szel jelöljük:

$$F(x) = \mathrm{P}\,(\,(-\infty, x]\,)$$

Az $F(x)$ függvény neve: **eloszlásfüggvény**. Az eloszlásfüggvény értékét egy x helyen úgy kapjuk meg, hogy az x -nél kisebb vagy egyenlő összes k helyeken vesszük a súlyfüggvény értékét, és ezeket az értékeket összeadjuk:

$$F(x) = \sum_{k:k \le x} p(k)$$

Ha a szóbanforgó eloszlás egy X valószínűségi változó eloszlása, akkor $F(x)$ jelentése nyilván annak a valószínűsége, hogy a véletlen X kisebb vagy egyenlő mint x:

$$F(x) = \mathrm{P}(X \le x)$$

Nyilvánvaló, hogy minden eloszlásfüggvény monoton növekedő

$$F(x-1) \le F(x)$$

és súlyfüggvény értéke egy adott x helyen egyenlő az eloszlásfüggvény megváltozásával:

$$p(x) = F(x) - F(x-1)$$

Az $F(b) - F(a)$ különbség pedig nem más, mint az a -nál nagyobb, de b -t még meg nem haladó számokból álló halmaz valószínűsége:

$$F(b) - F(a) = \sum_{k: a<k\leq b} p(k) = \mathrm{P}(a < X \leq b)$$

1. Példa: Fiatal házaspárok – eloszlásfüggvény. Az előző alfejeztben elképzeltük, hogy véletlenszerűen választunk egy fiatal házaspárt, és tekintjük a gyerekeik számát, mint a X valószínűségi változót. Ennek a valószínűségi változónak az eloszlását ott megadtuk egy táblázattal. Most a táblázatot kiegészítjük egy harmadik sorral, ami az eloszlásfüggvény értékeit tartalmazza a megadott x helyeken:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1
$F(x)$	0.2	0.6	0.9	1.0

Vegyük észre, hogyan képződik a harmadik sor a másodikból: minden elem egyenlő a felette álló sorban tőle balra lévő és a felette lévő elemek összegével.

2. Példa: Két dobókockával dobott számok összege – eloszlásfüggvény. Ha két dobókockával dobva a dobott számok összegét tekintjük, akkor az így kapott valószínűségi változó eloszlását megadó második sor alá a harmadik sorba odaírhatjuk az eloszlásfüggvény értékeit:

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

3. Példa: Két dobókockával dobott számok szorzata – eloszlásfüggvény. Ha két dobókockával dobva a dobott számok szorzatát tekintjük, akkor az így kapott valószínűségi változó eloszlását megadó második sor alá a harmadik sorba odaírhatjuk az eloszlásfüggvény értékeit:

x	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{16}{36}$	$\frac{17}{36}$	$\frac{19}{36}$	$\frac{23}{36}$	$\frac{25}{36}$	$\frac{26}{36}$	$\frac{28}{36}$	$\frac{30}{36}$	$\frac{32}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

2.4. Jobboldali eloszlásfüggvény

Egy adott normált eloszlás esetén egy x lehetséges értékkel kapcsolatban a $[x, \infty)$ intervallum valószínűségét $T(x)$ -szel jelöljük:

$$T(x) = \mathrm{P}(\ [x, \infty)\ )$$

A $T(x)$ függvény neve: **jobboldali eloszlásfüggvény**. A jobboldali eloszlásfüggvény értékét egy x helyen úgy kapjuk meg, hogy az x -nél nagyobb vagy egyenlő összes k helyeken vesszük a súlyfüggvény értékét, és ezeket az értékeket összeadjuk:

$$T(x) = \sum_{k: k \geq x} p(k)$$

Amikor valamilyen probléma kapcsán a most bevezetett $T(x)$ jobboldali eloszlásfüggvényt is és a korábban bevezetett $F(x)$ eloszlásfüggvényt is használjuk, akkor az $F(x)$ -et **baloldali eloszlásfüggvény**nek is nevezzük.

Ha a szóbanforgó eloszlás egy X valószínűségi változó eloszlása, akkor $T(x)$ jelentése nyilván annak a valószínűsége, hogy a véletlen X nagyobb vagy egyenlő mint x:

$$T(x) = \mathrm{P}(X \geq x)$$

Nyilvánvaló, hogy minden jobboldali eloszlásfüggvény monoton csökkenő

$$T(x-1) \geq T(x)$$

Nyilvánvaló, hogy minden x -re fennáll az alábbi egyszerű összefüggés:

$$F(x) + T(x) - p(x) = 1$$

1. Példa: Fiatal házaspárok – jobboldali eloszlásfüggvény. Egy véletlenszerűen választott fiatal házaspár gyerekeinek számával, mint valószínűségi változóval foglalkoztunk a fentiekben. Ennek a valószínűségi változónak az eloszlását és eloszlásfüggvényét megadtuk egy táblázattal. Most a táblázatot kiegészítjük egy negyedik sorral, ami a jobboldali eloszlásfüggvény értékeit tartalmazza a megadott x helyeken:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1
$F(x)$	0.2	0.6	0.9	1.0
$T(x)$	1.0	0.8	0.4	0.1

2. Példa: Két dobókockával dobott számok összege – jobboldali eloszlásfüggvény. Két dobókockával dobva a dobott számok összegét vizsgáltuk. Ennek a valószínűségi változónak az eloszlását és eloszlásfüggvényét megadtuk egy táblázattal. Most a táblázatot kiegészítjük egy negyedik sorral, ami a jobboldali eloszlásfüggvény értékeit tartalmazza a megadott x helyeken:

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$
$T(x)$	$\frac{36}{36}$	$\frac{35}{36}$	$\frac{33}{36}$	$\frac{30}{36}$	$\frac{26}{36}$	$\frac{21}{36}$	$\frac{15}{36}$	$\frac{10}{36}$	$\frac{6}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

3. Példa: Két dobókockával dobott számok szorzata – jobboldali eloszlásfüggvény. Két dobókockával dobva a dobott számok szorzatát vizsgáltuk. Ennek a valószínűségi változónak az eloszlását és eloszlásfüggvényét megadtuk egy táblázattal. Most a táblázatot kiegészítjük egy negyedik sorral, ami a jopbboldali eloszlásfüggvény értékeit tartalmazza a megadott x helyeken:

x	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{16}{36}$	$\frac{17}{36}$	$\frac{19}{36}$	$\frac{23}{36}$	$\frac{25}{36}$	$\frac{26}{36}$	$\frac{28}{36}$	$\frac{30}{36}$	$\frac{32}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$
$T(x)$	$\frac{36}{36}$	$\frac{35}{36}$	$\frac{33}{36}$	$\frac{31}{36}$	$\frac{28}{36}$	$\frac{26}{36}$	$\frac{22}{36}$	$\frac{20}{36}$	$\frac{19}{36}$	$\frac{17}{36}$	$\frac{13}{36}$	$\frac{11}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

2.5. Medián

Adott diszkrét valószínűségi változó vagy eloszlás esetén az x számot **medián**nak nevezzük, ha úgy osztja ketté a számegyenest, hogy a $(-\infty; x]$ intervallum is és a $[x; +\infty)$ intervallum is legalább $\frac{1}{2}$ valószínűségű:

$$\mathrm{P}\left(\, (-\infty; x] \,\right) \geq \frac{1}{2} \qquad \text{és} \qquad \mathrm{P}\left(\, [x; +\infty) \,\right) \geq \frac{1}{2}$$

azaz

$$F(x) \geq \frac{1}{2} \qquad \text{és} \qquad T(x) \geq \frac{1}{2}$$

1. Példa: Fiatal házaspárok gyermekei – medián. Egy véletlenszerűen választott fiatal házaspár gyerekeinek számával, mint valószínűségi változóval foglalkoztunk a fentiekben. Ennek a valószínűségi változónak a mediánja 1. Ezt a tényt az eloszlásfüggvény és a jobboldali eloszlásfüggvény

x	0	**1**	2	3
$p(x)$	0.2	0.4	0.3	0.1
$F(x)$	0.2	0.6	0.9	1.0
$T(x)$	1.0	0.8	0.4	0.1

táblázatából kiolvasható

$$F(1) = 0.6 \geq \frac{1}{2} \qquad \text{és} \qquad T(1) = 0.8 \geq \frac{1}{2}$$

egyenlőtlenségekből látjuk.

2. Példa: Két dobókockával dobott számok összege – medián. Két dobókockával dobva a dobott számok összegét vizsgáltuk. Ennek a valószínűségi változónak a mediánja 7. Ezt a tényt az eloszlásfüggvény és a jobboldali eloszlásfüggvény

x	2	3	4	5	6	**7**	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$
$T(x)$	$\frac{36}{36}$	$\frac{35}{36}$	$\frac{3}{36}$	$\frac{30}{36}$	$\frac{26}{36}$	$\frac{21}{36}$	$\frac{15}{36}$	$\frac{10}{36}$	$\frac{6}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

táblázatából kiolvasható

$$F(7) = \frac{21}{36} \geq \frac{1}{2} \qquad \text{és} \qquad T(7) = \frac{21}{36} \geq \frac{1}{2}$$

egyenlőtlenségekből látjuk.

3. Példa: Két dobókockával dobott számok szorzata – medián. Két dobókockával dobva a dobott számok szorzatát vizsgáltuk. Ennek a valószínűségi változónak a mediánja 10. Ezt a tényt az eloszlásfüggvény és a jobboldali eloszlásfüggvény

x	1	2	3	4	5	6	8	9	**10**	12	15	16	18	20	24	25	30	36
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{16}{36}$	$\frac{17}{36}$	$\frac{19}{36}$	$\frac{23}{36}$	$\frac{25}{36}$	$\frac{26}{36}$	$\frac{28}{36}$	$\frac{30}{36}$	$\frac{32}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$
$T(x)$	$\frac{36}{36}$	$\frac{35}{36}$	$\frac{33}{36}$	$\frac{31}{36}$	$\frac{28}{36}$	$\frac{26}{36}$	$\frac{22}{36}$	$\frac{20}{36}$	$\frac{19}{36}$	$\frac{17}{36}$	$\frac{13}{36}$	$\frac{11}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

táblázatából kiolvasható

$$F(10) = \frac{19}{36} \geq \frac{1}{2} \qquad \text{és} \qquad T(10) = \frac{19}{36} \geq \frac{1}{2}$$

egyenlőtlenségekből látjuk.

2.6. Kétdimenziós valószínűségi változó

1. Példa: Fiatal házaspárok – gyerekek és nagyszülők. Tegyük fel, hogy egy szociológiai felmérésben fiatal házaspárok (akik 3 évnél nem régebben házasodtak) gyermekeinek számát és az élő nagyszülők számát vizsgálták. A gyermekek száma 0, 1, 2, 3 lehet, a nagyszülők száma pedig 0, 1, 2, 3, 4. Ez összesen 4-szer 5, azaz 20 lehetőséget ad, melyeket egy táblázatba célszerű elrendezni. A tapasztalatból tudhatjuk a 20 lehetőség mindegyikének százalékos értékét. Tegyük fel, hogy ezek az alábbiak:

nagyszülők száma					
4	6.0	12.0	9.0	3.0	
3	8.0	16.0	12.0	4.0	
2	3.0	6.0	4.5	1.5	
1	2.0	4.0	3.0	1.0	
0	1.0	2.0	1.5	0.5	
	0	1	2	3	gyerekek száma

(Ellenőrizhető, hogy a 20 darab beírt százalék érték összege 100.)

Ha egy fiatal házaspárt véletlenszerűen választunk, és tekintjük az alábbi két valószínűségi változót:

$$X = \text{ gyerekek száma}, \quad Y = \text{ nagyszülők száma}$$

akkor az $(X;Y)$ kétdimenziós valószínűségi változó 20 lehetséges értéke mindegyikének a valószínűségét (a megfelelő százalék érték alapján) beírva a táblázatba megkapjuk az $(X;Y)$ kétdimenziós valószínűségi változó eloszlását:

y					
4	0.060	0.120	0.090	0.030	
3	0.080	0.160	0.120	0.040	
2	0.030	0.060	0.045	0.015	
1	0.020	0.040	0.030	0.010	
0	0.010	0.020	0.015	0.005	
	0	1	2	3	x

(Ellenőrizhető, hogy a 20 darab beírt valószínűség érték összege 1.)

2. Példa: Hány piros, hány kék – ismét. Az 1. fejezet 5. alfejezében az 5. példában feltettük, hogy egy dobozban 45 darab golyó van, melyek közül 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 darab golyót visszatevés nélkül, és megfigyeljük, hogy a kivett golyók között hány piros és hány kék van:

$$X = \text{ ahány piros van a kivett golyók között}$$

$$Y = \text{ ahány kék van a kivett golyók között}$$

Akkor ott kiszámoltuk annak a valószínűségét, hogy a kivett golyók között pontosan x darab piros és pontosan y darab kék lesz. A válasz ez volt:

$$\frac{\binom{10}{x}\binom{15}{y}\binom{20}{8-x-y}}{\binom{45}{8}}$$

minden olyan x -re és y -ra, melyek eleget tesznek a következő egyenlőtlenségeknek:

$$0 \le x \le 8$$

$$0 \le y \le 8$$

$$0 \le 8 - x - y \le 8$$

Ezekből a valószínűségekből az alábbi táblázatot raktuk ott össze:

y										
8	0.000									
7	0.001	0.000								
6	0.004	0.005	0.001							
5	0.016	0.026	0.013	0.002						
4	0.031	0.072	0.054	0.015	0.001					
3	0.033	0.102	0.108	0.048	0.009	0.001				
2	0.019	0.076	0.106	0.067	0.019	0.002	0.000			
1	0.005	0.027	0.049	0.040	0.017	0.003	0.000	0.000		
0	0.001	0.004	0.008	0.009	0.005	0.001	0.000	0.000	0.000	
	0	1	2	3	4	5	6	7	8	x

Vegyük észre, hogy a fenti képlet, illetve ez a táblázat nem más, mint az $(X;Y)$ kétdimenziós valószínűségi változó eloszlása matematikai képlettel, illetve táblázattal megadva..

Kérdés: Mi a valószínűsége annak, hogy $Y < 2X$, azaz a kihúzott kékek száma kevesebb mint a kihúzott pirosak számának a kétszerese?

Válasz: Ha előszedjük középiskolás tudásunkat, és meggondoljuk, hogy az $y < 2x$ egyenlőtlenség milyen $(x;y)$ -okra teljesül, akkor könnyen ellenőrizhetjük, hogy azokra az $(x;y)$ -okra teljesül, amilyen helyekre 1-eket tettünk az alábbi táblázatban:

y										
8	..0..	..0..	..0..	..0..	..0..	..0..	..1..	..1..	..1..	
7	..0..	..0..	..0..	..0..	..1..	..1..	..1..	..1..	..1..	
6	..0..	..0..	..0..	..0..	..1..	..1..	..1..	..1..	..1..	
5	..0..	..0..	..0..	..1..	..1..	..1..	..1..	..1..	..1..	
4	..0..	..0..	..0..	..1..	..1..	..1..	..1..	..1..	..1..	
3	..0..	..0..	..1..	..1..	..1..	..1..	..1..	..1..	..1..	
2	..0..	..0..	..1..	..1..	..1..	..1..	..1..	..1..	..1..	
1	..0..	..1..	..1..	..1..	..1..	..1..	..1..	..1..	..1..	
0	..0..	..1..	..1..	..1..	..1..	..1..	..1..	..1..	..1..	
	0	1	2	3	4	5	6	7	8	x

A kérdezett valószínűséget úgy kapjuk meg, hogy összeadjuk azokat a valószínűségeket az $(X;Y)$ eloszlásának a táblázatában, melyek az 1-eknek megyfelelő helyen vannak. Ezt az összeadást az Excelben a `SUMPRODUCT` (magyarul: `SZORZATÖSSZEG`) utasítással nagyon egyszerű végrehajtani.

2.7. Kétdimenziós eloszlás és súlyfüggvény

Ha egy kétdimenziós diszkrét valószínűségi változó minden lehetséges értékének megadjuk a valószínűségét (táblázattal, képlettel, stb.), akkor azt mondjuk, hogy megadjuk a **kétdimenziós valószínűségi változó eloszlását**.

Ha (mindenféle valószínűségi változóra való tekintet nélkül) egy véges vagy megszámlálhatóan végtelen síkbeli halmaz (x, y) elemeihez nemnegatív $p(x, y)$ számokat rendelünk úgy, hogy azok összege 1:

$$\sum_{(x,y)} p(x,y) = 1$$

akkor azt mondjuk, hogy megadunk egy **kétdimenziós (más szóval: síkbeli) valószínűségi eloszlás**t, más kifejezéssel kétdimenziós (más szóval: síkbeli) **normált eloszlás**t. A $p(x, y)$ függvényt **súlyfüggvény**nek (vagy **valószínűségi függvény**nek) nevezzük.

Egy (X, Y) kétdimenziós valószínűségi változó esetén $p(x, y)$ annak a valószínűségét adja meg, hogy a véletlen (X, Y) érték (x, y) -szel egyenlő, vagyis a véletlen X érték x -szel egyenlő, és a véletlen Y érték pedig y -nal egyenlő:

$$p(x,y) = \mathrm{P}(X = x, Y = y)$$

2.8. Módusz

Egy diszkrét valószínűségi változó lehetséges értékei közül a legvalószínűbbet a valószínűségi változó **módusz**ának nevezzük. Ha több ilyen érték is van, akkor több módusz is van. Ha egy eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk, akkor azt az x értéket, mely(ek)re $p(x)$ maximális, **az eloszlás módusza(i)**nak nevezzük.

1. Példa: Fiatal házaspárok gyermekei – módusz. Egy véletlenszerűen választott fiatal házaspár gyerekeinek számával, mint valószínűségi változóval foglalkoztunk a fentiekben. Ennek a valószínűségi változónak a módusza 1. Ezt a tényt a valószínűségi változó eloszlásának táblázatából kiolvashatjuk:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

2. Példa: Két dobókockával dobott számok összege – módusz. Két dobókockával dobva a dobott számok összegét vizsgáltuk. Ennek a valószínűségi változónak a módusza 7. Ezt a tényt a valószínűségi változó eloszlásának táblázatából kiolashatjuk:

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

3. Példa: Két dobókockával dobott számok szorzata – módusz. Két dobókockával dobva a dobott számok szorzatát vizsgáltuk. Ennek a valószínűségi változónak két módusza van: a 6 és a 12. Ezt a tényt a valószínűségi változó eloszlásának táblázatából kiolashatjuk:

x	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2.9. Valószínűségek megadása súlyokkal

Ócska a pénztárcám. Egyrészt kevés benne a papírpénz (ez talán nem a pénztárca hibája), másrészt – és most erre kell odafigyleni – az aprópénz kihullik belőle, és a táskám alját nyomja. Kis unokám nagyon élvezi, ha ott turkálhat. Véletlenszerűen választ és kivesz egy érmét, aztán tanulmányozza, nézegeti. Vagy visszateszi, vagy nem. Aztán ugyanezt teszi megint, megint, és így tovább. Tegyük fel, hogy a választáskor egy-egy érme esélye arányos az érme súlyával. Ez a feltevés vitatható, de eléggé elfogadhatónak tűnik. Hogy igazából mi a valószínűség legelfogadhatóbb numerikus értéke, az kis unokámat sem érdekli, és most itt minket sem.

Eléggé érzékeny mérleggel megmértem az érméket. A tizedgrammnyi pontossággal mért tömegek és – az elfogadott arányossági elv szerinti – valószínűségek így festenek:

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**		
tömeg (g)	4.2	6.1	7.0	7.7	8.0	9.5	42.5	tömegek összege
vsz	**0.10**	**0.14**	**0.17**	**0.18**	**0.19**	**0.22**	1.00	vsz-ek összege

Minden egyes érme valószínűségét úgy számoltuk ki, hogy a tömegét elosztottuk az érmék tömegeinek összegével. Ez az egyszerű művelet az arányokat megtartja, és garantálja azt, hogy a valószínűségek összege 1 legyen.

Más elvek szerint is felvehetjük a valószínűségeket. Lehet, hogy valakinek az a hipotézis tűnik elfogadhatóbbnak, hogy az érmék valószínűségei az átmérőikkel arányosak. Másvalaki azt gondolhatja, hogy – az érméket korongoknak tekintve – a korongok területei a meghatározóak a valószínűségeik szempontjából, és ezért a valószínűségeket a területekkel arányosnak tekinti. Tolómércével nem volt nehéz megmérni az érmék átmérőit és kiszámolni a korongok területét, majd pedig ezekből az adatokból a valószínűségeket meghatározni. Íme ezeknek az adatoknak a táblázata:

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**		
átmérő (mm)	21.2	24.5	26.3	27.5	23.8	28.3	151.6	átmérők összege
vsz	**0.14**	**0.16**	**0.17**	**0.18**	**0.16**	**0.19**	1.00	vsz-ek összege

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**		
terület (mm^2)	353.0	471.4	543.3	594.0	444.9	629.0	3035.5	területek összege
vsz	**0.12**	**0.15**	**0.18**	**0.19**	**0.15**	**0.21**	1.00	vsz-ek összege

Kicsit furcsa módon, de mégis így van: ha – mondjuk – azok az átmérők azok az adatok, melyekkel arányosan vesszük fel a valószínűségeket, akkor is **súlyok**nak nevezzük ezeket az adatokat. Tehát ilyenkor az átmérők adják a súlyokat. Ha pedig a korongok területeivel arányosan vesszük fel a valószínűségeket, akkor azt mondjuk, hogy a valószínűségek felvételéhez a korongok területei **adják a súlyokat**. Figyelem! A súly szó ilyen értelmű használata nem tévesztendő össze a valószínűségek eloszlását jellemző, korábban definiált súlyfüggvény fogalmával.

Mivel a 100 forintos érme súlyosabb, mint az 50 -es, de az 50 -es átmérője nagyobb, mint 100 forintosé, a valószínűségek szempontjából nem mindegy, hogy miket tekintünk súlyoknak. Ha a valószínűségeket az érmék tömege (súlya) alapján vesszük fel, akkor a 100 forintos érme valószínűbb, mint az 50 -es. Ha viszont az érmék átmérője vagy korongjuk területe alapján, akkor az 50 forintos érme valószínűbb, mint a 100 forintos.

Feladat: Érmét kotorászok a táska aljából. Táskám aljában 5 darab 50 forintos, 1 darab 100 forintos, 2 darab 200 forintos érme lapul. Az érmék súlyai:

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**
tömeg (g)	4.2	6.1	7.0	7.7	8.0	9.5

Feltételezzük, hogy az érmék valószínűségei arányosak a súlyaikkal. Mi a valószínűsége annak, hogy 100 forintost vagy 200 forintost húzunk?

Megoldás:

érme	**5 darab 50 Ftos**	**1 darab 100 Ftos**	**2 darab 200 Ft-os**		
	5 x 7.7 =	1 x 8.0 =	2 x 9.5 =		
tömeg (g)	38.3	8.0	19	65.3	súlyok összege
vsz	**0.59**	**0.12**	**0.29**	1.00	vsz-ek öszege

A kérdezett valószínűség: $\mathbf{0.12} + \mathbf{0.29} = \mathbf{0.41}$, ami természetesen úgy is megkaphattunk volna, hogy a **0.59** valószínűséget 1 -ből kivonjuk.

2.10. Gyakorló feladatok

- Az alábbi 4 feladatban az ott definiált valószínűségi változók eloszlását adja meg táblázattal!
- Készítsen ábrákat az eloszlásokról Excellel!
- Adja meg táblázattal a bal- és jobb oldali eloszlásfüggvényt!
- Készítsen ábrákat az eloszlásfüggvényekről is!
- Keresse meg az eloszlások móduszát és mediánját is! (A számolásokhoz használhatja az Excelt.)
- Adja meg az eloszlásokat matemaikai képlettel is!

Jótanács: Egy X valószínűségi változó eloszlásának meghatározása kissé összetett feladat: meg kell keresni a valószínűségi változó lehetséges értékeit, és minden lehetséges értékeknek meg kell határozni a valószínűségét. A P$(X = x)$ vagy P$(X = k)$ (van amikor az x, van amikor a k betűt szeretjük használni) általános képlet megtalálása nehéz lehet. Ilyenkor érdemes lépésről lépésre haladni:

- Először a P$(X = 1)$ valószínűség értékére keressük meg a választ egy numerikus képlettel, és ennek örülünk.
- Ezután a P$(X = 2)$ valószínűség értékét adjuk meg egy numerikus képlettel, és ennek mégjobban örülünk.
- Most már nagyobb önbizalommal merünk belevágni a P$(X = 3)$ valószínűség értékének meghatározásába, és – ha még ez is sikerül, — akkor már nagyon örülünk.
- És így tovább lépésről lépésre haladva egyre jobban kivilágosodik előttünk, hogy mi a problémának a lényege, és – jó esetben – rájövünk még az általános képletre is!

Uccu neki, itt a lehetőség a módszer gyakorlására:

1. (a) Két szabályos érmével dobunk.
   (b) Három szabályos érmével dobunk.
   (c) Négy szabályos érmével dobunk.

(d) Öt szabályos érmével dobunk.
(e) Tíz szabályos érmével dobunk.

Mindegyik esetben a vizsgálandó valószínűségi változó:

$$X = \text{ ahány fejet kapunk}$$

2. (a) Addig dobunk egy szabályos érmével, amíg végre fejet kapunk.
   (b) Addig dobunk egy szabályos érmével, amíg másodszorra fejet kapunk.
   (c) Addig dobunk egy szabályos érmével, amíg harmadszorra fejet kapunk.

   Mindegyik esetben a vizsgálandó valószínűségi változó:

$$X = \text{ ahány dobás kell ehhez}$$

3. (a) Addig dobunk egy szabályos érmével, amíg végre fejet kapunk.
   (b) Addig dobunk szabályos érmével, amíg másodszorra fejet kapunk.
   (c) Addig dobunk szabályos érmével, amíg harmadszorra fejet kapunk.

   Mindegyik esetben a vizsgálandó valószínűségi változó:

$$X = \text{ ahány írás adódik eközben}$$

4. (a) Addig dobunk egy szabályos dobókockával, amíg végre hatost kapunk,
   (b) Addig dobunk egy szabályos dobókockával, amíg másodszorra hatost kapunk,
   (c) Addig dobunk egy szabályos dobókockával, amíg harmadszorra hatost kapunk,

   Mindegyik esetben a vizsgálandó valószínűségi változó:

$$X = \text{ ahány dobás kell ehhez}$$

5. 45 darab golyó van egy dobozban. Közülük 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 golyót **visszatevés nélkül**. Legyen X a pirosak, Y a kékek száma a kihúzottak között! Az $(X;Y)$ kétdimenziós valószínűségi változó eloszlása valahol korábban megtalálható ebben a jegyzetben.
   (a) Keresse meg!
   (b) Állítsa elő Excellel ezt a táblázatot!
   (c) Adja meg az X valószínűségi változó eloszlását képlettel, táblázattal és grafikonnal!
   (d) Adja meg az Y valószínűségi változó eloszlását képlettel, táblázattal és grafikonnal!

6. 45 darab golyó van egy dobozban. Közülük 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 golyót, de most nem visszatevés nélkül, hanem **visszatevéssel**. Legyen X a pirosak, Y a kékek száma a kihúzottak között!
   (a) Adja meg az $(X;Y)$ kétdimenziós valószínűségi változó eloszlását képlettel is!
   (b) Állítsa elő Excellel az eloszlás táblázatát!
   (c) Adja meg az X valószínűségi változó eloszlását képlettel, táblázattal és grafikonnal!
   (d) Adja meg az Y valószínűségi változó eloszlását képlettel, táblázattal és grafikonnal!

7. Az érmék súlyai:

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**
tömeg (g)	4.2	6.1	7.0	7.7	8.0	9.5

Feltételezzük, hogy az érmék valószínűségei arányosak a súlyaikkal. Hány darabot tegyünk be az egyes érmékből egy kalapba, hogy a kalapból való húzáskor az 5 forintos érme legyen a legvalószínűbb, aztán a 10 forintos, és így tovább, a legkevésbé valószínűszínű a 200 forintos legyen! Törekedjen arra, hogy minimális számú érmével oldja meg a feladatot! A megoldásnál – ha gondolja – használjon számítógépet!

3. Folytonos egyenletes eloszlás

Bár a könyvnek az első részében diszkrét valószínűségi változókkal foglalkozunk, ebben a fejezetben folytonos valószínűségi változókkal fogunk megismerkedni. Egyelőre csak a legegyszerűbb folytonos problémákat ismerjük meg. Bonyolultabb folytonos problémákat a könyv későbbi részeiben fogunk tárgyalni.

3.1. Folytonos egyenletes eloszlás

1. Egyenletes eloszlás intervallumon Tekintsünk egy véges, pozitív hosszúságú I intervallumot. Tegyük fel, hogy valaki az intervallumban választ egy véletlen pontot, vagyis a lehetséges pontok az I intervallum pontjai, azaz az eseménytér az I intervallum. Ha az I intervallum bármely részintervalluma esetén annak a valószínűsége, hogy a véletlen pont a részintervallumba esik

$$\textit{részintervallum valószínűsége} = \frac{\textit{részintervallum hossza}}{\textit{I hossza}}$$

akkor azt mondjuk, hogy a véletlen pont egyenletes eloszlású az I intervallumon. Ilyenkor egy részintervallum valószínűsége csak a részintervallum hosszától függ, de attól, hogy maga a részintervallum, hol helyezkedik el az eseménytéren belül, attól nem függ. Ha egy részintervallumot eltolunk az eseménytéren belül, akkor ettől a részintervallum valószínűsége nem változik meg. Ez e két tény indokolja az "egyenletes eloszlás" elnevezést.

2. Egyenletes eloszlás síkbeli halmazon: Tekintsünk egy véges, pozitív területű S halmazt a síkon. Tegyük fel, hogy valaki a halmazban választ egy véletlen pontot, vagyis a lehetséges pontok az S halmaz pontjai, azaz az eseménytér az S halmaz. Ha az S halmaz bármely részhalmaza esetén annak a valószínűsége, hogy a véletlen pont a részhalmazba esik

$$\textit{részhalmaz valószínűsége} = \frac{\textit{részhalmaz területe}}{\textit{S területe}}$$

akkor azt mondjuk, hogy a véletlen pont egyenletes eloszlású az S halmazon. Ilyenkor egy részhalmaz valószínűsége csak a részhalmaz területétől függ, de attól, hogy maga a részhalmaz, hol helyezkedik el az eseménytéren belül, attól nem függ. Ha egy részhalmazt eltolunk az eseménytéren belül, akkor ettől a részhalmaz valószínűsége nem változik meg.

3. Egyenletes eloszlás térbeli halmazon: Anélkül, hogy részleteznénk, megemlítjük, hogy egy véges, pozitív térfogatú S térrészen vett egyenletes eloszlás esetén:

$$\textit{részhalmaz valószínűsége} = \frac{\textit{részhalmaz térfogata}}{\textit{S térfogata}}$$

Tekinteve, hogy a hosszúság, a terület, a térfogat számítása általában a geometria körébe tartozik, az ilyen valószínűségszámítási problémákat **geometriai problémák**nak is nevezzük.

Egyenletes eloszlású, független koordináták. Megemlítjük azt a nyilvánvaló tényt, hogy ha egy síkbeli, illetve térbeli pont koordinátáit egymástól függetlenül választjuk meg egy-egy intervallumban, akkor a pont egyenletes eloszlású lesz az intervallumok által (direktszorzatként) meghatározott téglalapban, illetve vagy téglatestben.

Feladat: Átrepül-e a teniszlabda a kerítésünkön? Kerítésünk 3 cm átmérőjű függőleges, hosszú rudakból áll. A rudak 20 cm periódussal ismétlődnek, a köztük lévő rések 17 cm-esek. A kerítésnek háttal állva, néhány méterről, merőlegesen nekidobok a kerítésnek egy 5 cm átmérőjű teniszlabdát. Mi a valószínűsége, hogy a teniszlabda a vasrudak érintése nélkül átrepül közöttük?

Megoldás: Amikor a teniszlabda eléri vagy elérné az oszlopok síkját, a középpontnak a tőle balra legközelebbi rúd középvonalától való távolsága véletlentől függ, jelöljük a cm-ekben vett távolságot X -szel, így X lehetséges értékei a $[0; 20]$ intervallumot teszik ki, és X nyilván egyenletes eloszlást követ. Az érintés nélküli átrepülés feltétele, hogy az $4 < X < 16$ egyenlőtlenségek teljesüljenek. Ennek valószínűsége:

$$\text{P(érintés nélküli átrepül)} = \frac{16-4}{20} = \frac{12}{20} = 0.6$$

Feladat: Átrepül-e a teniszlabda a szomszéd kerítésén? A szomszéd kerítése ugyanilyen függőleges rudakbkól áll, de az ő kerítésében vízszintes rudak is vannak 30 cm-es periódussal. Mi a valószínűsége, hogy a szomszéd kerítésén átrepül a teniszlabda?

1. Megoldás: Amikor a teniszlabda eléri vagy elérné a kerítés síkját, a középpontnak a tőle balra legközelebbi rúd középvonalától való távolsága legyen X, az alatta lévő vízszintes rúd középvonalától való távolsága legyen Y. Az (X, Y) lehetséges értékei a $[0; 20]$ és a $[0; 30]$ intervallumok által meghatározott téglalapot teszik ki, és (X, Y) nyilván egyenletes eloszlást követ. Az érintés nélküli átrepülés feltétele, hogy az $4 < X < 16$, $4 < Y < 26$ egyenlőtlenségek teljesüljenek, vagyis (X, Y) egy kisebb téglalapban legyen. Ennek valószínűsége a két téglalap területének a hányadosa:

$$\text{P(érintés nélkül átrepül)} = \frac{(16-4)\cdot(26-4)}{20\cdot 30} = \frac{12\cdot 22}{600} = 0.44$$

Megjegyzés: X és Y függetlensége miatt így is okoskodhattunk volna:

$$\text{P(érintés nélkül átrepül)} =$$

$$= \text{P}(4 < X < 16, 4 < Y < 26) = \text{P}(4 < X < 16)\cdot\text{P}(4 < Y < 26) = \frac{12}{20}\cdot\frac{22}{30} = 0.44$$

Feladat: Utazás busszal, metróval. Reggelente busszal és metróval megyek az egyetemre, és az átszállás közben még a reggelimet is megveszem. A busz 10 percenként jár, a metró 5 percenként. Mivel az induláskor nem taktikázok, a megállóban a buszra való X várakozási időm egyenletes eloszlású 0 és 10 perc között. Kiszámíthatlan, hogy a reggeli vásárlása hogyan alakul, ezért a metró állomáson a várakozással eltöltött Y időm egyenletes eloszlású 0 és 5 perc között, akármennyi is az X értéke .

- Mi a valószínűsége, hogy a metróra többet kell várnom, mint a buszra?
- Mi a valószínűsége, hogy a várakozással eltöltött összes időm több, mint 4 perc?

Megoldás:

- Az (X, Y) lehetséges értékei a $[0; 10]$ és a $[0; 5]$ intervallumok által meghatározott téglalapot teszik ki, és ezen a téglalapon (X, Y) egyenletes eloszlást követ. Az $Y > X$ esemény ebben a téglalapban egy háromszöget jelöl ki, melynek területe a téglalap területének a negyede. Ezért

$$\text{P(a metróra többet kell várnom, mint a buszra)} = \frac{1}{4} = 0.25$$

- Az $X + Y > 4$ esemény a téglalapban egy ötszöget határoz meg, melynek komplementere egy derékszögű háromszög. A derékszögű háromszög területe 8 terület egység, ezért az ötszögé 50 - 8 = 42 terület egység. A keresett valószínűség:

$$\text{P(a várakozással eltöltött összes időm több, mint 4 perc)} = \frac{42}{50} == 0.84$$

Feladat: Randevú. Jancsi és Juliska déli 12 óra és 1 óra között randevúznak. Azonban a randevú menetébe és sikerébe beleszól a véletlen is: az 1 órás időtartam alatt egymástól független pillanatban érkeznek, mindketten egyenletes eloszlás szerint, és – megbeszélésük szerint – 20 percet várakoznak, aztán elmennek. Így aztán vagy találkoznak, vagy nem. Mi a valószínűsége, hogy találkoznak?

1. Megoldás: Jancsi érkezési pillanatát jelöljük X -szel, Juliskáét Y -nal. Mivel X és Y függetlenek és egyenletes eloszlásúak 0 és 1 között, az (X, Y) pont egyenletes eloszlású a $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$ pontok által meghatározott négyzetben. Nyilvánvaló, hogy a találkozó létrejöttének feltétele, hogy az

$$Y \geq X - 1/3 \quad , \quad Y \leq X + 1/3$$

egyenlőtlenségek teljesüljenek, vagyis az (X, Y) pontnak az

$$y = x - 1/3$$

egyenletű egyenes és az

$$y = x + 1/3$$

egyenletű egyenes közötti tartományban kell lenni. Ez a tartomány egy hatszög. A hatszög komplementere a négyzetben két derékszögű háromszög, melyeknek befogyói $2/3$ hosszúak. A két háromszögből egy kis négyzetet lehet összerakni, melynek oldalhossza $2/3$. Ezért a két háromszög együttes területe $4/9$, hatszögé pedig $1 - 4/9 = 5/9$. A keresett valószínűség egyenlő a hatszög területe osztva a négyzet területével (ami r1), ezért:

$$\text{P(találkoznak)} = \frac{5}{9} = 0.5556 = 0.57$$

2. Megoldás: Az előző megoldásban – helyesen – folytonos modellt használtunk, most – kissé pontatlan, de tanulságos – diszkrét modellt adunk a problémára: az időt egész percekben fogjuk mérni. Ilyen szemlélet mellett a fiatalok érkezési pillanatai (jelöljük ezeket most is X -szel és Y -nal) egyenletes eloszlást követnek az $\{1, 2, \ldots, 59, 60\}$ almazon, és az (X, Y) számpár egyenletes eloszlást követ a 3600 elemből álló halmazon, melyet az alábbi ábrán szemléltetünk:

	1	2	3	4	5	6	.	.	.	55	56	57	58	59	60	X
1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
4	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
6	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
.	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
.	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
.	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
55	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
56	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
57	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
58	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
59	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
60	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
Y																

A találkozó létrejöttének feltétele most az, hogy az

$$Y \geq X - 20 \quad , \quad Y \leq X + 20$$

egyenlőtlenségek teljesüljenek, vagyis az $(X; Y)$ pontnak az az alábbi ábrán $*$ -gal jelölt pontok halmazába kell esni.

	1	.	.	.	20	21	.	.	.	40	41	.	.	.	60	X
1	$*$	$*$	$*$	$*$	$*$	$*$										
.	$*$	$*$	$*$	$*$	$*$	$*$	$*$									
.	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$								
.	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$							
20	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$						
21	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$					
.		$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$				
.			$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$			
.				$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$		
40					$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	
41						$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	
.							$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	
.								$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	
.									$*$	$*$	$*$	$*$	$*$	$*$	$*$	
60										$*$	$*$	$*$	$*$	$*$	$*$	
Y																

Egyszerű elemi feladat megszámolni, hogy hány $*$ található ezen az ábrán: 2040. Tehát a találkozó valószínűségére – ezzel a modellel – az alábbi eredményt kapjuk:

$$\text{P(találkoznak)} = \frac{2\,040}{3\,600} = 0.5667 = 0.57$$

Az eredményt összevetve az 1. megoldás eredményével látjuk, hogy a diszkrét modell kb. 0.01 -dal nagyobb valószínűséget ad, mint a folytonos modell. Ez a hiba nem nagy, de jelzi, hogy oda kell figyelni arra, hogy folytonos problémát folytonos modellel kezeljünk.

3. Megoldás: Ha percek helyett másodpeercekkel dolgozunk a modellben, akkor az összes esetek száma 3 600 -nak a négyzete, ami 12 960 000, a kedvező esetek számára pedig 7 202 400 adódik, amiből a

$$\text{P(találkoznak)} = \frac{7\,202\,400}{12\,960\,000} = 0.5557$$

valószínűséget kapjuk. Ezt az eredményt összevetve az 1. megoldás eredményével látjuk, hogy ez a diszkrét modell már csak kb. 0.001 -del nagyobb valószínűséget ad mint a folytonos modell.

Feladat: Buffon féle tű probléma. Párhuzamos egyeneseket húzunk egy nagy papírra (vagy a földre) egymástól 20 cm távolságra. Egy 10 cm hosszú tűt elég magasról hetykén leejtünk. A tű vagy metszi valamelyik egyenest, vagy egyiket sem metszi. Mi a valószínűsége annak, hogy a tű metszi valamelyik egyenest?

Megoldás: Jelöljük X -szel a tű által meghatározott egyenesnek a párhuzamos egyenesekkel bezárt hegyes szögét radiánban mérve, Y -nal pedig a tű középpontjának a hozzá legközelebb lévő egyenestől való távolságát. Nyilván $0 \leq X \leq \pi/2$, illetve $0 \leq Y \leq 10$. X és Y függetlensége miatt (X, Y) egyenletes eloszlású a $[\,0\,;\,\pi/2\,]$, illetve $[\,0\,;\,10\,]$ intervallumok által meghatározott 5π területű T téglalapon. Egyszerű trigonometriai probléma annak ellenőrzése, hogy a metszés pontosan akkor áll fenn, ha $Y < 5\sin(X)$, azaz az (X, Y) pont az

$$y = 5\sin(x)$$

egyenletű görbe alatti A tartományba esik. Ezért a keresett valószínűség:

$$\text{P(metszés)} = \frac{A \text{ területe}}{T \text{ területe}} = \frac{\int_0^{\pi/2} 5\sin(x)\,dx}{5\pi} = \frac{\int_0^{\pi/2} \sin(x)\,dx}{\pi} = \frac{1}{\pi}$$

3.2. RAND utasítás

Az Excelben a 0 és 1 közötti értékeket felvevő

```
RAND(),  magyarul VÉL()
```

utasítás olyan véletlen számot állít elő, mely egyenletes eloszlást követ a $[0;1]$ intervallumon. Az üres zárójelpár nem elírás: az Excel formai szabályai szerint a `RAND` mögé oda kell írni a `()` üres zárójelpárt. A `RAND` utasítással generált véletlen számokat **random szám**oknak hívjuk. Ha valaki a `RANDBETWEEN` utasítással generált véletlen számokat is **random szám**oknak hívja, akkor illik utalni rá, hogy 0 és 1 közötti, vagy egész értékű-e a véletlen szám.

Jelölés: Jegyzetünkben a `RAND` utasítás által előállított véletlen szám jelölésére RND -t írunk. Több véletlen szám használata esetén azokat – a matematika szokásai szerint – indexezéssel különböztetjük meg egymástól: az RND_1 és RND_2 jelölésekben az indexek arra utalnak, hogy két különböző véletlen számról van szó. Az Excel nem használ indexeket, az Excelben `RAND` utasítás többszöri alkalmazása különböző, független véletlen számokat jelentenek. Például a

```
2*RAND() + 3*RAND()
```

Excel utasításra jegyzetünkben az indexeket is tartalmazó

$$2\,\mathrm{RND}_1 + 3\,\mathrm{RND}_2$$

képlettel utalunk. Ha indexek nélkül

$$2\,\mathrm{RND} + 3\,\mathrm{RND}$$

írnánk, akkor matematikai

$$2\,\mathrm{RND} + 3\,\mathrm{RND} \;=\; 5\,\mathrm{RND}$$

összevonás miatt a

```
2*RAND() + 3*RAND()
```

Excel utasítást helytelenül összekeverhetnénk a

```
5*RAND()
```

utasítással.

3.3. Random számok tulajdonságai

A `RAND` utasítást az Excelben úgy találták ki, hogy a random számokra igazak az alábbiak:

1. Akármilyen x szám esetén annak a valószínűsége, hogy egy random szám pontosan egyenlő x -szel, nulla:

$$\mathrm{P}(\mathrm{RND} = x) = 0$$

2. $0 \le a \le b \le 1$ esetén annak a valószínűsége, hogy egy random szám az a és b által meghatározott intervallumba esik, egyenlő az intervallum hosszával. Az, hogy az intervallum zárt, nyitott vagy félig zárt, félig nyitott, közömbös:

$$\mathrm{P}(a \le \mathrm{RND} \le b) = b - a$$
$$\mathrm{P}(a \le \mathrm{RND} < b) = b - a$$
$$\mathrm{P}(a < \mathrm{RND} \le b) = b - a$$
$$\mathrm{P}(a < \mathrm{RND} < b) = b - a$$

3. Bármely x számra, ami 0 és 1 között van, igaz, hogy

$$\mathrm{P}(\mathrm{RND} \le x) = \mathrm{P}(\mathrm{RND} < x) = x$$

4. A random számok fontos tulajdonsága, hogy ha két random számból, mint koordinátákból egy számpárt rakunk össze, akkor az $(\mathrm{RND}_1, \mathrm{RND}_2)$ pont egyenletes eloszlást követ az egységnyi oldalú négyzetben. Ez – többek között – azt is jelenti, hogy ha A a négyzetnek akármilyen részhalmaza, akkor

$$\mathrm{P}(\ (\mathrm{RND}_1, \mathrm{RND}_2) \in A\ ) = A \text{ területe}$$

5. Ha három random számból, mint koordinátákból egy $(\mathrm{RND}_1, \mathrm{RND}_2, \mathrm{RND}_3)$ pontot rakunk össze a 3 -dimenziós térben, akkor ez a pont egyenletes eloszlást követ a tér egységnyi oldalú kockájában. Ez – többek között – azt is jelenti, hogy ha A a kockának akármilyen részhalmaza, akkor

$$(\mathrm{RND}_1, \mathrm{RND}_2, \mathrm{RND}_3) \in A\ ) = \text{ az } A \text{ halmaz térfogata}$$

6. Ha n random számból, mint koordinátákból egy $(\mathrm{RND}_1, \mathrm{RND}_2, \dots, \mathrm{RND}_n)$ pontot rakunk össze az n -dimenziós térben, akkor ez a pont egyenletes eloszlást követ az n -dimenziós tér egységnyi oldalú kockájában. Ez – többek között – azt is jelenti, hogy ha A az n -dimenziós kockának akármilyen részhalmaza, akkor

$$\mathrm{P}(\ (\mathrm{RND}_1, \mathrm{RND}_2, \dots, \mathrm{RND}_n) \in A\ ) = \text{ az } A \text{ halmaz } n\text{-dimenziós térfogata}$$

7. Megjegyezzük, hogy azok a tulajdonságok, melyeket az előző három pontban fogalmaztunk meg, igazából azt jelentik, hogy ha több random számot állítunk elő Excellel, akkor azoknak egymáshoz semmi közük sincsen, azok egymástól függetlenek. A függetlenség matematikai definícióját később tanuljuk.

3.4. Lineáris transzformációk

1. **Nyújtás**
   Ha az RND random számot
   megszorzunk egy 1-nél nagyobb a számmal,
   akkor egy új valószínűségi változót kapunk, melyet $a\,\mathrm{RND}$ -vel szokás jelölni. Nyilvánvaló, hogy
   a RND egyenletes eloszlást követ a $[0; a]$ intervallumon.

2. **Zsugorítás**
   Ha az RND random számot
   megszorzunk egy 1-nél kisebb pozitív a számmal,
   akkor egy új valószínűségi változót kapunk, melyet $a\,\mathrm{RND}$ -vel szokás jelölni. Nyilvánvaló, hogy
   a RND egyenletes eloszlást követ a $[0; a]$ intervallumon.

3. **Eltolás**
   Ha az RND random számhoz
   hozzáadunk egy b számot,
   akkor egy új valószínűségi változót kapunk, melyet $(\mathrm{RND} + b)$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(RND + b)$ egyenletes eloszlást követ a $[b; b + 1]$ intervallumon.

4. **Nyújtás és eltolás (vagy: zsugorítás és eltolás)**
   Ha az RND random számot
   megszorzunk egy pozitív a számmal, és a szorzathoz hozzáadunk egy b számot,
   akkor egy új valószínűségi változót kapunk, melyet $(a\,\mathrm{RND} + b)$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(a\,RND + b)$ egyenletes eloszlást követ a $[b; a + b]$ intervallumon.

5. **Tükrözés az origóra**
   Ha az RND random számnak
   vesszük az ellentettjét (vagyis megszorozzunk (-1) -gyel),
   akkor egy új valószínűségi változót kapunk, melyet $(-\mathrm{RND})$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(-RND)$ egyenletes eloszlást követ a $[-1; 0]$ intervallumon.

6. **Tükrözés és nyújtás**
   Ha az RND random számot
   megszorzunk egy 1-nél nagyobb abszolút értékű negatív a számmal,
   akkor egy új valószínűségi változót kapunk, melyet $(a\,\text{RND})$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(a\,RND)$ *egyenletes eloszlást követ az* $[a;0]$ *intervallumon* $(a<0)$.

7. **Tükrözés és zsugorítás**
   Ha az RND random számot
   megszorzunk egy egy 1-nél kisebb abszolút értékű negatív a számmal,
   akkor egy új valószínűségi változót kapunk, melyet $(a\,\text{RND})$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(a\,RND)$ *egyenletes eloszlást követ az* $[a;0]$ *intervallumon* $(a<0)$.

8. **Tükrözés, nyújtás és eltolás**
   Ha az RND random számot
   megszorzunk egy 1-nél nagyobb abszolút értékű negatív a számmal, és a szorzathoz hozzáadunk egy b számot,
   akkor egy új valószínűségi változót kapunk, melyet $(a\,\text{RND}+b)$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(a\,RND+b)$ *egyenletes eloszlást követ a* $[a+b;b]$ *intervallumon.*

9. **Tükrözés, zsugorítás és eltolás**
   Ha az RND random számot
   megszorzunk egy 1-nél kisebb abszolút értékű negatív a számmal, és a szorzathoz hozzáadunk egy b számot,
   akkor egy új valószínűségi változót kapunk, melyet $(a\,\text{RND}+b)$ -vel szokás jelölni. Nyilvánvaló, hogy
   $(a\,RND+b)$ *egyenletes eloszlást követ a* $[a+b;b]$ *intervallumon* $(a<0)$.

3.5. Gyakorló feladatok

1. Egy városban a metró szabályosan 5 percenként jár. Az én érkezési pillanatom a metróállomásra véletlenszerű. Attól függ, hogy hogyan ébredek, mennyi ideig vacakolok, hogyan tudok átmenni a zebrákon, stb. Vegyük a várakozási időmet egyenletes eloszlásúnak 0 és 5 perc között! Mi a valószínűsége annak, hogy
   (a) kevesebb, mint 3 percet kell várnom a metróra?
   (b) több, mint 3 percet kell várnom a metróra?
   (c) matematikai pontossággal pontosan 3 percet kell várnom a metróra?
   (d) kevesebb, mint x percet kell várnom a metróra, ha $0<x<5$?
   (e) kevesebb, mint x percet kell várnom a metróra, ha $x<0$?
   (f) kevesebb, mint x percet kell várnom a metróra, ha $5<x$?
   (g) matematikai pontossággal pontosan x percet kell várnom a metróra?

2. A reggeli buszom egyenletes eloszlású időpontban érkezik 7:30 és 7:40 között. Hányra menjek oda a megállóba, ha legalább 80% -os esélyt szeretnék arra, hogy elérjem?

3. Egyenletes eloszlás szerint választunk egy számot a $[-1;2]$ intervallumban. Jelöljük ezt X -szel. Mi annak a valószínűsége, hogy $X^3<0.5$?

4. Szimulálja Excellel az előző feladatban szereplő X -et , és sok kísérlet kapcsán számolja ki az ott szereplő esemény relatív gyakoriságát! Ha mindent jól csinált, akkor a relatív gyakoriságnak közel kell lenni a valószínűséghez.

5. Egy téglalapot úgy szerkesztünk meg, hogy mindkét oldalának hosszát egymástól függetlenül 0 és 10 cm között egyenletes eloszlás szerint választjuk. Mi a valószínűsége annak, hogy
   (a) a téglalap kerülete nagyobb 15 cm -nél?

(b) a területe kisebb 25 cm^2 -nél?

(c) a téglalap kerülete nagyobb 15 cm-nél, és a területe kisebb 25 cm^2 területegységnél?

6. Egy véletlen téglalapot úgy szerkesztünk meg, hogy mindkét oldalának hosszát egymástól függetlenül 0 és 1 között egyenletes eloszlás szerint választjuk. Mi a valószínűsége annak, hogy

    (a) a téglalap kerülete kisebb x hosszegységnél, ahol $0 < x < 4$?

    (b) a területe kisebb y területegységnél, ahol $0 < y < 1$?

    (c) a téglalap kerüete kisebb x hosszegységnél, és a területe kisebb y területegységnél?

7. Egy nagy papírlapra 5 cm-enként párhuzamos egyeneseket húzunk, majd egy 10 cm hosszú tűt magasról a papírra ejtünk. Mi a valószínűsége annak, hogy

    (a) a tű metszi valamelyik egyenest?

    (b) a tű két egyenest metsz?

8. *Bertrand-paradoxon (híres probléma!)*: Egyezzünk meg abban, hogy a kör egy húrját "hosszúnak" nevezzük, ha a húrhoz tartozó középponti szög 120 foknál nagyobb, vagyis a húr hosszabb, mint a körbe rajzolható egyenlőoldalú háromszög oldalának a hossza. Egységsugarú kör esetén ez annyit jelent, hogy a húr hosszabb, mint $\sqrt{3}$ egység. Mi a valószínűsége annak, hogy a kör húrjai közül véletlenszerűen választva hosszú húr adódik, ha a véletlenszerű választás az alábbi módszerek egyikét jelenti?

    (a) A kör egyik átmérőjét véletlenszerűen kiválasztjuk úgy, hogy az átmérő irányát kijelölő szög egyenletes eloszlású legyen 0 és 2π között, majd pedig a kiválasztott átmérőn egyenletes eloszlás szerint választunk egy pontot. Azt a húrt tekintjük, mely átmegy ezen a ponton, és merőleges az átmérőre.

    (b) A kör kerületén egymástól függetlenül két pontot választunk egyenletes eloszlás szerint, és tekintjük a két pont által meghatározott húrt.

    (c) A körlapon egyenletes eloszlás szerint választunk egy pontot, és tekintjük azt a húrt, aminek ez a pont a felezőpontja.

9. *A Buffon féle tű probléma általánosításai*:

    (a) Egy nagy papírlapra 10 cm-enként párhuzamos egyeneseket húzunk, majd egy 10 cm hosszú tűt magasról a papírra ejtünk. Mi a valószínűsége annak, hogy

        i. a tű metszi valamelyik egyenest?

        ii. metszi valamelyik egyenest, és 30 foknál kisebb szöget zár be az egyenessel?

    (b) Egy nagy papírlapra 5 cm-enként párhuzamos egyeneseket húzunk, majd egy 10 cm hosszú tűt magasról a papírra ejtünk. Mi a valószínűsége annak, hogy

        i. a tű metszi valamelyik egyenest?

        ii. a tű két egyenest metsz?

    (c) ***Extra feladat:***
    Egy nagy papírlapra 2 cm-enként párhuzamos egyeneseket húzunk, majd egy 10 cm hosszú tűt magasról a papírra ejtünk. Mi a valószínűsége annak, hogy a tű pontosan k darab egyenest metsz? ($k \geq 0$ egész)

    (d) ***Extra feladat:***
    Egy nagy papírlapra négyzethálót szerkesztünk úgy, hogy 20 cm-enként párhuzamos piros egyeneseket húzunk, majd az ezekre az egyenesekre merőlegesen szintén 20 cm-enként párhuzamos zöld egyeneseket húzunk. Ezután egy 10 cm hosszú tűt magasról a papírra ejtünk. Mi a valószínűsége annak, hogy

        i. a tű piros és zöld egyenest is metsz?

        ii. a tű metszi valamelyik egyenest?

10. ***Extra feladat:***
A korábban tálalt "Randevú probléma" folytatása. Jancsi és Juliska déli 12 óra és 1 óra között – mint feljebb leírtuk – randevúznak. Azonban a randevú menetébe és sikerébe nem csak a véletlen szól bele, mint korábban, hanem Juliska apukája is: az 1 órás időtartam alatt a fiatalok érkezésétől függetlenül ő is egyenletes eloszlás szerint odatoppan a randevú helyszínére, és ha valamelyik fiatalt ott találja, akkor nagy patáliát csap, és megakadályozza a randevú létrejöttét. Mi a valószínűsége, hogy a fiatalok randevúja mégis létrejön?

4. További műveletek és szabályok

4.1. Műveletek eseményekre

Események különbsége: Az A bekövetkezik, de a B nem. Vagyis A -nak és B -nek a különbsége nem más, mint $A \cap \overline{B}$. A különbség jelölése: $A \backslash B$. Tehát $A \backslash B = A \cap \overline{B}$.

Teljes eseményrendszer: Véges vagy végtelen sok egymást kizáró eseményeknek egy olyan rendszere, hogy ezeknek az eseményeknek az úniója a biztos esemény. Halmazelméleti nyelven mondva: az eseménytér "partíciója", azaz felbontása egymást kizáró halmazok úniójára.

Események növekvő sorozata: A sorozat bármely eleme maga után vonja a későbbieket:

$$A_1 \subset A_2 \subset A_3 \subset A_4 \subset \ldots$$

Események csökkenő sorozata: A sorozat bármely eleme maga után vonja a korábbiakat:

$$A_1 \supset A_2 \supset A_3 \supset A_4 \supset \ldots$$

Egy csökkenő sorozat esetén nyilván fennállnak az alábbiak:

$$A_1 \cap A_2 = A_2$$

$$A_1 \cap A_2 \cap A_3 = A_3$$

$$A_1 \cap A_2 \cap A_3 \cap A_4 = A_4$$

$$A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 = A_5$$

4.2. Szabályok eseményekre

Komplementer szabály: Bármely esemény komlementerének komplementere nem más, mint az eredeti esemény.

A sok egyéb, hasonlóan triviális szabály felsorolásától eltekintünk. Azonban – fontossága és furcsasága miatt – felhívjuk még a figyelmet a De Morgan szabályokra:

De Morgan szabály az únióra: Események úniójának komplementere egyenlő a komplementereik metszetével:

$$\overline{A_1 \cup A_2 \ldots \cup A_n} = \overline{A_1} \cap \overline{A_2} \ldots \cap \overline{A_n}$$

De Morgan szabály a metszetre: Események metszetének komplementere egyenlő a komplementereik úniójával:

$$\overline{A_1 \cap A_2 \ldots \cap A_n} = \overline{A_1} \cup \overline{A_2} \ldots \cup \overline{A_n}$$

4.3. Eloszlás transzformációja

Az új eloszlás valószínűségeinek kiszámolása összegzéssel történik. Ki kell gondolni, hogy az új eloszlás egyes tagjainak valószínűségeit a régi eloszlás mely tagjainak összegeként kapjuk meg.

1. Példa: Hány tagú a nagycsalád? Korábban vizsgáltuk az a problémát, melyben egy fiatal házaspárt véletlenszerűen választottunk, és tekintettük az alábbi két valószínűségi változót:

$$X = \text{ gyerekek száma}, \quad Y = \text{ nagyszülők száma}$$

Az akkor mondott feltételek mellett meghatároztuk az (X, Y) kétdimenziós valószínűségi változó eloszlását:

y					
4	0.060	0.120	0.090	0.030	
3	0.080	0.160	0.120	0.040	
2	0.030	0.060	0.045	0.015	
1	0.020	0.040	0.030	0.010	
0	0.010	0.020	0.015	0.005	
	0	1	2	3	x

Nézzük, hány tagú a nagycsalád, amikor a szülők, gyerekek mellett a nagyszülők is a családdal vannak. A nagycsalád létszáma nyilván $Z = 2 + X + Y$. Ha a fiatal házaspárt véletlenszerűen választjuk akkor X is, Y is valószínűségi változó. A Z eloszlásának mindenegyes tagját az (X, Y) kétdimenziós valószínűségi változó eloszlásából a megfelelő valószínűségek összegeként kapjuk meg:

$$\begin{aligned}
&\mathrm{P}(Z = 2) = 0.010 \\
&\mathrm{P}(Z = 3) = 0.040 = 0.020 + 0.020 \\
&\mathrm{P}(Z = 4) = 0.085 = 0.030 + 0.040 + 0.015 \\
&\mathrm{P}(Z = 5) = 0.175 = 0.080 + 0.060 + 0.030 + 0.005 \\
&\mathrm{P}(Z = 6) = 0.275 = 0.060 + 0.160 + 0.045 + 0.010 \\
&\mathrm{P}(Z = 7) = 0.255 = 0.120 + 0.120 + 0.015 \\
&\mathrm{P}(Z = 8) = 0.130 = 0.090 + 0.040 \\
&\mathrm{P}(Z = 9) = 0.030
\end{aligned}$$

Ha Z lehetséges értékeit és a hozzájuk tapadó valószínűségeket egy táblázatba rendezzük, megkapjuk Z eloszlását. Ezzel az eloszlással kell modelleznünk azt a problémát, ha egy véletlenszerűen választott fiatal házaspárhoz tartozó nagycsalád létszámát akarjuk vizsgálni.

z	2	3	4	5	6	7	8	9
$\mathrm{P}(Z = z)$	0.010	0.040	0.085	0.175	0.275	0.255	0.130	0.030

2. Példa: Három színes (piros vagy kék) golyó valószínűsége. Az 1. fejezet 7. alfejezében az 5. példában feltettük, hogy egy dobozban 45 darab golyó van, melyek közül 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 darab golyót visszatevés nélkül, és megfigyeljük, hogy a kivett golyók között hány piros és hány kék van:

$$X = \text{ ahány piros van a kivett golyók között}$$

$$Y = \text{ ahány kék van a kivett golyók között}$$

Tekintsük a $Z = X + Y$ valószínűségi változót, ami azt fejezi ki, hogy hány színes (piros vagy kék) golyó van a kihúzott 8 golyó között. **Kérdés:** Mi a valószínűsége annak, hogy $Z = 3$?

Első megoldás: Az (X, Y) eloszlását megadó

y										
8	0.000									
7	0.001	0.000								
6	0.004	0.005	0.001							
5	0.016	0.026	0.013	0.002						
4	0.0317	0.072	0.054	0.015	0.001					
3	0.033	0.102	0.108	0.048	0.009	0.001				
2	0.019	0.0756	0.106	0.067	0.019	0.002	0.000			
1	0.005	0.027	0.049	0.040	0.017	0.003	0.000	0.000		
0	0.001	0.004	0.008	0.009	0.005	0.001	0.000	0.000	0.000	
	0	1	2	3	4	5	6	7	8	x

táblázatban azokat a cellákat, melyek olyan (x, y) értkeknek felelnek meg, melyekre $x + y = 3$, és összadjuk a cellákban található valószínűségeket:

$$\mathrm{P}(Z = 3) = 0.0327 + 0.0755 + 0.0486 + 0.0086 = 0.1654$$

Második megoldás: Mivel a dobozban a 45 golyó között 25 színes golyó van, és közülük 8-at húzunk ki, a 3 színes golyó valószínűsége:

$$\frac{\binom{25}{3}\binom{45-25}{8-3}}{\binom{45}{8}} = 0.1654$$

3. Példa: Színes (piros vagy kék) golyók. (Az előző példa folytatása.)

Feladat: Határozzuk meg a Z valószínűségi változó eloszlását!

Első megoldás: Az (X, Y) eloszlását megadó fenti táblázatban Z minden lehetséges értékével kapcsolatban egy összeg adja meg a keresett valószínűséget:

$$\begin{array}{lcl}
\mathrm{P}(Z=0) = 0.001 & = & 0.001 \\
\mathrm{P}(Z=1) = 0.005 + 0.004 & = & 0.009 \\
\mathrm{P}(Z=2) = 0.019 + 0.027 + 0.008 & = & 0.054 \\
\mathrm{P}(Z=3) = 0.033 + 0.076 + 0.049 + 0.009 & = & 0.165 \\
\mathrm{P}(Z=4) = 0.031 + 0.102 + 0.106 + 0.040 + 0.005 & = & 0.284 \\
\mathrm{P}(Z=5) = 0.016 + 0.072 + 0.108 + 0.067 + 0.017 + 0.001 & = & 0.281 \\
\mathrm{P}(Z=6) = 0.004 + 0.026 + 0.054 + 0.048 + 0.019 + 0.003 + 0.000 & = & 0.156 \\
\mathrm{P}(Z=7) = 0.001 + 0.005 + 0.013 + 0.015 + 0.009 + 0.002 + 0.000 + 0.000 & = & 0.047 \\
\mathrm{P}(Z=8) = 0.000 + 0.000 + 0.001 + 0.002 + 0.001 + 0.001 + 0.000 + 0.000 + 0.000 & = & 0.005
\end{array}$$

Második megoldás: Mivel a dobozban a 45 golyó között 25 színes golyó van, és közülük 8-at húzunk ki, a z darab színes golyó valószínűsége:

$$\frac{\binom{25}{z}\binom{45-25}{8-z}}{\binom{45}{8}} \qquad (0 \le z \le 8)$$

Ha z helyére behelyettesítjük a $0, 1, 2, 3, 4, 5, 6, 7, 8$ értékeket, ugyanazt az eredményt kapjuk, mint az első megoldásban.

4.4. Síkbeli eloszlás vetületei

Az új eloszlás valószínűségeinek kiszámolása összegzéssel történik. Ki kell gondolni, hogy az új eloszlás egyes tagjainak valószínűségeit a régi eloszlás mely tagjainak összegeként kapjuk meg.

1. Példa: Fiatal házaspárok – vetítés. Tegyük fel, hogy az (X, Y) kétdimenziós valószínűségi változó eloszlása az alábbi táblázatban adott eloszlás:

y					
4	0.060	0.120	0.090	0.030	
3	0.080	0.160	0.120	0.040	
2	0.030	0.060	0.045	0.015	
1	0.020	0.040	0.030	0.010	
0	0.010	0.020	0.015	0.005	
	0	1	2	3	x

Könnyen ellenőrizhető, hogy az X eloszlásának mindenegyes tagját, vagyis a

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

táblázatban álló valószínűségeket az (X, Y) eloszlásából a megfelelő oszlopban található valószínűségek összegeként kapjuk meg:

P(X=1) = 0.060 + 0.080 + 0.030 + 0.020 + 0.010 = 0.2
P(X=2) = 0.120 + 0.160 + 0.060 + 0.040 + 0.020 = 0.4
P(X=3) = 0.090 + 0.120 + 0.045 + 0.030 + 0.015 = 0.3
P(X=4) = 0.030 + 0.040 + 0.015 + 0.010 + 0.005 = 0.1

Könnyű látni, hogy az Y eloszlásának mindenegyes tagját az (X, Y) eloszlásából a megfelelő sorban található valószínűségek összegeként lehet megkapni:

P(Y=4) = 0.060+ 0.120+ 0.090+ 0.030 = 0.30
P(Y=3) = 0.080+ 0.160+ 0.120+ 0.040 = 0.40
P(Y=2) = 0.030+ 0.060+ 0.045+ 0.015 = 0.15
P(Y=1) = 0.020+ 0.040+ 0.030+ 0.010 = 0.10
P(Y=0) = 0.010+ 0.020+ 0.015+ 0.005 = 0.05

2. Példa: Piros és kék golyók – vetítés. Az 1. fejezet 7. alfejezében az 5. példában feltettük, hogy egy dobozban 45 darab golyó van, melyek közül 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 darab golyót visszatevés nélkül, és megfigyeljük, hogy a kivett golyók között hány piros és hány kék van:

$$X = \text{ ahány piros van a kivett golyók között}$$

$$Y = \text{ ahány kék van a kivett golyók között}$$

Ott meghatároztuk az (X, Y) kétdimenziós valószínűségi változó eloszlását, egy táblázatot kaptunk. Gondoljuk meg ennek a példának a kapcsán is, hogy hogyan lehet (X, Y) eloszlásából X, illetve Y eloszlását meghatározni!

Első megoldás: Az (X, Y) kétdimenziós valószínűségi változó eloszlását jelentő táblázat oszlopainak összegzésével kapjuk az X eloszlását:

$$
\begin{array}{lll}
\mathrm{P}(X=0) = & 0.000+0.001+0.004+0.016+0.031+0.033+0.019+0.005+0.001 & =0.109 \\
\mathrm{P}(X=1) = & 0.000+0.005+0.026+0.072+0.102+0.076+0.027+0.004 & =0.312 \\
\mathrm{P}(X=2) = & 0.001+0.013+0.054+0.108+0.106+0.049+0.008 & =0.339 \\
\mathrm{P}(X=3) = & 0.002+0.015+0.048+0.067+0.040+0.009 & =0.181 \\
\mathrm{P}(X=4) = & 0.001+0.009+0.019+0.017+0.005 & =0.051 \\
\mathrm{P}(X=5) = & 0.001+0.002+0.003+0.001 & =0.008 \\
\mathrm{P}(X=6) = & 0.000+0.000+0.000 & =0.001 \\
\mathrm{P}(X=7) = & 0.000+0.000 & =0.000 \\
\mathrm{P}(X=8) = & 0.000 & =0.000
\end{array}
$$

Teljesen hasonló módon – a táblázat sorainak összegzésével – kapjuk az Y eloszlását:

$$
\begin{array}{lll}
\mathrm{P}(Y=8) = & 0.000 & = 0.000 \\
\mathrm{P}(Y=7) = & 0.001+0.000 & = 0.001 \\
\mathrm{P}(Y=6) = & 0.004+0.005+0.001 & = 0.010 \\
\mathrm{P}(Y=5) = & 0.016+0.026+0.013+0.002 & = 0.057 \\
\mathrm{P}(Y=4) = & 0.031+0.072+0.054+0.015+0.001 & = 0.174 \\
\mathrm{P}(Y=3) = & 0.033+0.102+0.108+0.048+0.009+0.001 & = 0.301 \\
\mathrm{P}(Y=2) = & 0.019+0.076+0.106+0.067+0.019+0.002+0.000 & = 0.289 \\
\mathrm{P}(Y=1) = & 0.005+0.027+0.049+0.040+0.017+0.003+0.000+0.000 & = 0.142 \\
\mathrm{P}(Y=0) = & 0.001+0.004+0.008+0.009+0.005+0.001+0.000+0.000+0.000 & = 0.027
\end{array}
$$

Második megoldás: Az X eloszlásának meghatározása közvetlenül is történhet, hiszen ha a dobozban a 45 golyó között 10 piros golyó van, és közülük 8-at húzunk ki, akkor az x darab piros golyó valószínűsége:

$$
\frac{\binom{10}{x}\binom{45-10}{8-x}}{\binom{45}{8}} \qquad (0 \leq x \leq 8)
$$

Ha x helyére behelyettesítjük a $0, 1, 2, 3, 4, 5, 6, 7, 8$ értékeket, ugyanazt az eredményt kapjuk, mint az első megoldásban.

Az Y eloszlásának meghatározása is hasonlóan történhet, hiszen ha a dobozban a 45 golyó között 15 kék golyó van, és közülük 8-at húzunk ki, akkor az y darab kék golyó valószínűsége:

$$
\frac{\binom{15}{y}\binom{45-15}{8-y}}{\binom{45}{8}} \qquad (0 \leq y \leq 8)
$$

Ha y helyére behelyettesítjük a $0, 1, 2, 3, 4, 5, 6, 7, 8$ értékeket, ugyanazt az eredményt kapjuk, mint az első megoldásban.

4.5. Szabályok valószínűségekre

1. **Összegzési szabály három tetszőleges eseményre**: Ha A, B, C tetszőleges események, akkor

$$
\begin{array}{rcl}
\mathrm{P}(A \cup B \cup C) & = & \\
& & \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) \\
& - & \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) \\
& + & \mathrm{P}(A \cap B \cap C)
\end{array}
$$

Megjegyzés. Vegyük észre, hogy az egyenlőség jobb oldalán

- az első sorban $\binom{3}{1} = 3$ tag áll, az egyes események valószínűségei $+$ jelekkel
- a második sorban $\binom{3}{2} = 3$ tag áll, az eseményekből alkotható párok metszeteinek valószínűségei $-$ jelekkel
- a harmadik sorban $\binom{3}{3} = 1$ tag áll, a három esemény metszetének valószínűsége $+$ jellel

2. **Összegzési szabály több (tetszőleges) eseményre – avagy "Poincaré" vagy "szita" formula**: Ha $A_1, A_2, \dots, A_n$ tetszőleges események, akkor

$$
\begin{aligned}
\mathrm{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = & \\
& \mathrm{P}(A_1) + \mathrm{P}(A_2) + \ldots + \mathrm{P}(A_n) \\
- & \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \ldots - \mathrm{P}(A_{n-1} \cap A_n) \\
+ & \mathrm{P}(A_1 \cap A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_4) + \ldots + \mathrm{P}(A_{n-2} \cap A_{n-1} \cap A_n) \\
\vdots & \\
+ & (-1)^{n+1}\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_n)
\end{aligned}
$$

Megjegyzés. Vegyük észre, hogy az egyenlőség jobb oldalán

- az első sorban $\binom{n}{1} = n$ tag áll, az egyes események valószínűségei $+$ jelekkel
- a második sorban $\binom{n}{2}$ tag áll, az eseményekből alkotható párok metszeteinek valószínűségei $-$ jelekkel
- a harmadik sorban $\binom{n}{3}$ tag áll, az eseményekből alkotható hármasok metszeteinek valószínűségei $+$ jelekkel
- az utolsó sorban $\binom{n}{n} = 1$ tag áll, az összes esemény metszetének valószínűsége $+$ vagy $-$ jellel attól függően, hogy n páratlan vagy páros

3. **Folytonossági szabály események növekvő sorozatára** *(Extra tananyag)*: Ha végtelen sok esemény növekvő sorozatot alkot, akkor úniójuk valószínűsége egyenlő az események valószínűségeinek határértékével:

$$\mathrm{P}(A_1 \cup A_2 \cup \ldots) = \lim_{n\to\infty} \mathrm{P}(A_n)$$

4. **Folytonossági szabály események csökkenő sorozatára** *(Extra tananyag)*: : Ha végtelen sok esemény csökkenő sorozatot alkot, akkor metszetük valószínűsége egyenlő az események valószínűségeinek határértékével:

$$\mathrm{P}(A_1 \cap A_2 \cap \ldots) = \lim_{n\to\infty} \mathrm{P}(A_n)$$

Feladat: Annak esélye, hogy mindenki hűtlenkedik. *(Extra tananyag)*: Egy mulatságon, melyen 10 házaspár rock and roll számokra táncol, azt eszelik ki, hogy a változatosság kedvéért minden szám előtt kisorsolják, hogy ki kivel táncoljon. Minden férj nevét cédulára írják. Minden szám előtt a cédulákat kalapba teszik, minden feleség kihúz egy cédulát, és a soron következő számot azzal táncolja, akinek a nevét kihúzta. Vajon mi a valószínűsége annak, hogy minden feleség "hűtlenkedik", azaz nem a saját férjével táncol?

Megoldás: Aki elsőnek húz, az 10 cédula közül választ. Aki másodiknak húz, az 9 cédula közül választ. És így tovább, aki utolsónak húz, az csak 1 cédula közül válszt. Ezért a 10 hölgy és a 10 férfi

$$10 \cdot 9 \cdot \ldots \cdot 1 \; = \; 10\,!$$

darab – nyilván egyformán valószínű – felállásban táncolhat. Tehát egy klasszikus problémával van dolgunk, ahol az elemi események a lehetséges felállások a táncparketten.

Definiáljuk az $A_1, A_2, A_3, \ldots, A_{10}$ eseményeket a következőképpen: A_1 az az esemény, hogy a legidősebb feleség a férjével táncol, A_2 az az az esemény, hogy a második legidősebb feleség a férjével táncol, és így tovább. A feladat a

$$\text{minden feleség hűtlenkedik} \;=\; \overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_{10}}$$

esemény valószínűségét kérdezi. A De Morgan azonosság felhasználásával

$$\mathrm{P}\left(\overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_{10}} \right) \;=\; 1 - \mathrm{P}\left(A_1 \cup A_2 \cup \ldots \cup A_{10} \right)$$

A $\mathrm{P}(A_1 \cup A_2 \cup \ldots \cup A_{10})$ valószínűséget a Poincaré formula felhasználásával fogjuk meghatározni. A Poincaré formula jobb oldalán szereplő tagok közül először kiragadjuk a

$$\mathrm{P}(A_1 \cap A_2 \cap A_3)$$

tagot, és a benne szereplő eseményt jellemezzük. A $A_1 \cap A_2 \cap A_3$ esemény jelentése: a legidősebb, a második legidősebb, és a harmadik legidősebb feleség a saját férjével táncol, a többi feleség pedig a többi 7 férj akármelyikével. Ez $7\,!$ lehetőséget jelent, ezért

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \frac{7\,!}{10\,!}$$

Hasonlóképpen

$$\mathrm{P}(A_1 \cap A_2 \cap A_4) = \frac{7\,!}{10\,!}$$

$$\vdots$$

$$\mathrm{P}(A_{n-2} \cap A_{n-1} \cap A_n) = \frac{7\,!}{10\,!}$$

A Poincaré formula jobb oldalának 3 -ik sora ezeknek a tagoknak az összege:

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_4) + \ldots + \mathrm{P}(A_{n-2} \cap A_{n-1} \cap A_n)$$

Mivel az összegnek minden tagja egyenlő $\frac{7\,!}{10\,!}$ -sal, és a tagok száma $\binom{10}{7}$, az összeg értéke

$$\binom{10}{7} \cdot \frac{7\,!}{10\,!} \;=\; \frac{1}{3\,!}$$

A Poincaré formula jobb oldalának r -ik sora hasonlóan meghatározható. Az értéke

$$\binom{10}{r} \cdot \frac{(10-r)\,!}{10\,!} \;=\; \frac{1}{r\,!}$$

A Poincaré formula jobb oldala ezeknek a valószínégeknek a váltakozó előjellel vett összege. Ezért

$$\mathrm{P}\left(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_{10} \right) =$$

$$= \frac{1}{1\,!} - \frac{1}{2\,!} + \frac{1}{3\,!} - \ldots - \frac{1}{10\,!}$$

A keresett valószínűséget a komplementer szabállyal kapjuk:

$$\mathrm{P}\left(\text{minden feleség hűtlenkedik} \right) \;=\; 1 - \mathrm{P}\left(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_{10} \right) =$$

$$= 1 - \frac{1}{1\,!} + \frac{1}{2\,!} - \frac{1}{3\,!} + \ldots + \frac{1}{10\,!}$$

1. Megjegyzés: A megoldás gondolatmenetéből világos, hogy ha nem 10, hanem n házaspár van a mulatságon, akkor

$$\mathrm{P}\left(\text{minden feleség hűtlenkedik} \right) \;=\; 1 - \frac{1}{1\,!} + \frac{1}{2\,!} - \frac{1}{3\,!} + \ldots + (-1)^n \frac{1}{n\,!}$$

Mint – a Taylor sorok elméletéből – jól ismert, ennek a kifejezésnek a határérétke $n \to \infty$ esetén e^{-1}, ahol $\mathrm{e} = 2.71\ldots$ a természetes logaritmus alapja. Ezért sok házaspár esetén a

$$\mathrm{P}\,(\text{ minden feleség hűtlenkedik }) \approx \frac{1}{\mathrm{e}} \quad (\approx 0.37\,)$$

2. Megjegyzés: Közösségekben gyakran sorsolnak – például karácsonykor – abból a célból, hogy ki kit ajándékozzon meg. Mindenki felírja a nevét egy-egy cédulára, a cédulákat kalapba teszik, aztán mindenki húz egy cédulát, hogy a kihúzott embernek adjon majd ajándékot. Meg szoktak lepődni az emberek, amikor valaki saját magát húzza! Pedig egyáltalán nem kellene ezen meglelődni, hiszen annak az esélye, hogy senki sem húzza ki önmagát (azaz mindenki "hűtlenkedik") – mint kiszámoltuk – körülbelül 0.37, a komplementer eseményé, azaz hogy legalább egy ember önmagát húzza, ennél lényegesen nagyobb, $1 - 0.37 = 0.63$.

4.6. Gyakorló feladatok

1. (a) Mi a valószínűsége annak, hogy 10 szabályos dobókockával dobva a legnagyobb kapott szám kisebb vagy egyenlő, mint 4, azaz mind a 10 kapott szám kisebb vagy egyenlő, mint 4?
   (b) Mi a valószínűsége annak, hogy 10 szabályos dobókockával dobva a legnagyobb kapott szám kisebb vagy egyenlő,, mint 3?
   (c) Mi a valószínűsége annak, hogy 10 szabályos dobókockával dobva a legnagyobb kapott szám 4 -gyel egyenlő?
2. Ha egy szabályos dobókockát végtelen sokszor feldobnánk, akkor minden n -re értelmezhető lenne az alábbi esemény: "*az első n dobás során nem dobunk hatost*".
   (a) Mennyi ennek az eseménynek a valószínűsége?
   (b) Győződjön meg róla, hogy ezek az események csökkenő sorozatot alkotnak!
   (c) Mit jelent ennek a végtelen sok eseményeknek a metszete?
   (d) A fentiekből kiadódik annak az eseménynek a valószínűsége, hogy "*a végtelen sok dobás során soha sem dobunk 6 -ost*". Rakja össze fentiekből, hogy hogyan jön ez ki!
3. Venn diagram segítségével igazolja az összegzési szabályt két tetszőleges eseményre: ha A, B tetszőleges események, akkor
$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$
4. Venn diagram segítségével igazolja az összegzési szabályt három tetszőleges eseményre: ha A, B, C tetszőleges események, akkor
$$\begin{aligned}\mathrm{P}(A \cup B \cup C) \quad = & \\ & \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) \\ - & \ \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) \\ + & \ \mathrm{P}(A \cap B \cap C)\end{aligned}$$
5. Hogyan nézhet ki az $(X;Y)$ kétdimenziós valószínűségi változó eloszlása, hogy X és Y között fennáll a
   (a) $Y = 20 - X$
   (b) $Y = 2\,X$
   (c) $Y \leq 20 - X$
   (d) $Y = X^2$

(e) $Y < X^2$

relációk valamelyike?

6. Adjon meg olyan síkbeli egyenletes eloszlást, melynek vetületei egyik tengelyen sem egyenletes!
7. Adjon meg olyan síkbeli nem egyenletes eloszlást, melynek vetülete mindkét tengelyen egyenletes!
8. Adjon meg két egymástól különböző síkbeli eloszlást, melyeknek vetületei mindkét tengelyen megegyeznek! (Tanulság: egy síkbeli eloszlás vetületei nem határozzák meg a síkbeli eloszlást.)

5. Feltételes valószínűség és eloszlás

5.1. Feltételes valószínűség

Legyenek A és B események valamely véletlen jelenséggel kapcsolatban. Képzeljük el, hogy N kísérletet végzünk a jelenségre. Jelöljük N_A -val, hogy hányszor következik be az A esemény, és jelöljük $N_{A\cap B}$ -vel, hogy hányszor következik be az A -val együtt a B is. Másképpen mondva: N_A az A esemény gyakorisága, $N_{A\cap B}$ az $A\cap B$ esemény gyakorisága. A következő hányadost **feltételes relatív gyakorság**-nak nevezzük:

$$\frac{N_{A\cap B}}{N_A}$$

Részletesebben kifejezve a hányados neve: **a** B **eseménynek az** A **eseményre vonatkozó feltételes relatív gyakorsága**. A tört értéke azt mutatja, hogy azok között az esetek között, amikor A bekövetkezik, hányad részben, milyen arányban következik be A -val együtt a B is.

A számlálót is és a nevezőt is N -nel osztva, azt kapjuk, hogy

$$\frac{N_{A\cap B}}{N_A} = \frac{\frac{N_{A\cap B}}{N}}{\frac{N_A}{N}} \approx \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}$$

vagyis – sok kísérlet esetén – a feltételes relatív gyakorság körülbelül egyenlő az alábbi kifejezéssel:

$$\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}$$

Ezt az értéket nevezzük **a** B **esemény feltételes valószínűségének, feltéve, hogy** A **bekövetkezik**. A feltételes valószínűséget $\mathrm{P}(B|A)$ -vel jelöljük:

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}$$

Ezt a formulát a **valószínűségek osztási szabály**ának is nevezzük. Ha $\mathrm{P}(A) = 0$, akkor a a hányados nem definiált. Ilyenkor a feltételes valószínűség ezzel a hányadossal nem értelmezhető.

1. Megjegyzés: Ha B maga után vonja A -t, vagyis $B \subset A$, akkor $A\cap B = B$. Ilyenkor az osztási szabály így egyszerűsödik:

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(B)}{\mathrm{P}(A)} \qquad \text{ha} \quad B\subset A$$

2. Megjegyzés: Hasonlóképpen értelmezhető az A feltételes valószínűsége, feltéve, hogy a B bekövetkezik:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$$

Ha A maga után vonja B -t, vagyis $A\subset B$, akkor

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A)}{\mathrm{P}(B)} \qquad \text{ha} \quad A\subset B$$

3. Megjegyzés: A $\mathrm{P}(A|B)$ és a $\mathrm{P}(B|A)$ feltételes valószínűségek lehetnek egymással egyenlők is, de általában nem egyenlőek. Például ha A azt jelenti, hogy a szabályos dobókockával 5 -nél kisebbet dobok, B azt, hogy 3 -nál nagyobbat, akkor $\mathrm{P}(\,B\,|\,A\,) = \frac{1}{4}$, illetve $\mathrm{P}(\,A\,|\,B\,) = \frac{1}{3}$. Világos, hogy $\mathrm{P}(\,A\,|\,B\,) = \mathrm{P}(\,B|\,A)$ akkor és csak akkor, ha $\mathrm{P}(A) = \mathrm{P}(B)$.

4. Megjegyzés: Nyilvánvaló, hogy ha $A = S$, vagyis A a biztos esemény, akkor P$(B|S) =$ P(B).

5. Megjegyzés: Az is nyilvánvaló, hogy ha A és B kizáró események, akkor P$(B|A) = 0$ és P$(A|B) = 0$.

6. Megjegyzés: Az is nyilvánvaló, hogy ha $A \subset B$, vagyis A maga után vonja B -t, akkor P$(B|A) = 1$.

Feladat: Ha van fiú, van-e lány is? Ha egy véletlenszerűen választott kétgyerekes családban van fiú, akkor mi a valószínűsége annak, hogy lány is van? (Feltételezzük, hogy minden gyerek, függetlenül a többi gyerektől, 0.5 valószínűséggel születik fiúnak, 0.5 valószínűséggel lánynak.)

Megoldás:

$$\text{P(van lány | van fiú)} = \frac{\text{P(van fiú és van lány)}}{\text{P(van fiú)}} = \frac{\left(\frac{2}{4}\right)}{\left(\frac{3}{4}\right)} = \frac{2}{3}$$

1 . Megjegyzés: Egy véletlenszerűen választott kétgyerekes család gyerekeit – statisztikai szempontból – helyettesíthetjük két szabályos érmével: mondjuk a "fej" jelentsen "fiú"-t, az írás jelentsen "lány"-t. Dobjuk fel a két érmét sokszor (100-szor, 200-szor, még többször), és azon esetek között, amikor van fej (azaz van fiú a családban), megnézhetjük, hogy hányadrészben van írás is (azaz van lány a családban). Az arány körülbelül 2/3 lesz.

2. Megjegyzés: Ha az érméket számítógéppel szimuláljuk, akkor a nagy számú kísérlet elvégzése sem jelent gondot. Hajrá!

2. Feladat: Barátom tényleg beteg? Tegyük fel, hogy egy bizonyos ritka betegségben az embereknek csupán 1 ezreléke szenved. A betegséget egy olyan vizsgálattal lehet kimutatni, ami sajnos mindkét irányban tévedhet: beteg emberek esetén csak 0.8 a valószínűsége annak, hogy a vizsgálat jelzi a betegséget, egészséges emberek esetében pedig csak 0.9 a valószínűsége annak, hogy az egészséges embert egészségesnek jelzi. Barátomat nemrég vizsgálták, és a vizsgálat betegnek jelezte. Ő nagyon megijedt. Nyugtassuk meg barátomat, hogy nem kell megijednie: a vizsgálat eredményének ismeretében sem túl valószínű, hogy a kérdéses betegségben szenved.

Megoldás:

$$\begin{aligned}
&\text{P(beteg | betegnek diagnosztizálják)}\\
&= \frac{\text{P(beteg és betegnek diagnosztizálják)}}{\text{P(betegnek diagnosztizálják)}}\\
&= \frac{\text{P(beteg)}\cdot\text{P(betegnek diagnosztizálják | beteg)}}{\text{P(beteg)}\cdot\text{P(betegnek diagnosztizálják | beteg)}+\text{P(nem beteg)}\cdot\text{P(betegnek diagnosztizálják | nem beteg)}}\\
&= \frac{0.001 \cdot 0.8}{0.001 \cdot 0.8 + 0.999 \cdot 0.1}\\
&= 0.008
\end{aligned}$$

Látjuk, hogy annak az esélye, hogy a barátom beteg, ilyen információk mellett csak 8 ezred, ami még 1 százaléknál is kisebb, tehát egyáltalán nem sok.

Megjegyzés: Ilyen rosszul működő teszt esetén – ha nincs más lehetőség, és megengedhető, akkor – természetes ötlet, hogy több viszágalatot hajtsunk végre, és azok eredményéből következtessünk a helyzetre. Ezt az ötletet később, a függetlenség fogalmának bevezetése után fogjuk elemezni.

5.2. Szorzási szabályok

Szorzási szabály két eseményre:

$$P(A \cap B) = P(A)\ P(B|A)$$

Tehát annak a valószínűségét, hogy két esemény mindegyike bekövetkezik úgy lehet kiszámolni, hogy

- vesszük az első esemény valószínűségét,
- és szorzunk annak a valószínűségével, hogy a második bekövetkezik, feltéve, hogy az első bekövetkezett.

Szorzási szabály három eseményre:

$$P(A \cap B \cap C) = P(A)\ P(B|A)\ P(C|A \cap B)$$

Tehát annak a valószínűségét, hogy három esemény mindegyike bekövetkezik úgy lehet kiszámolni, hogy

- vesszük az első esemény valószínűségét,
- ezt megszorozzuk annak a valószínűségével, hogy a második bekövetkezik, feltéve, hogy az első bekövetkezett,
- és szorzunk annak a valószínűségével, hogy a harmadik bekövetkezik, feltéve, hogy az első és a második bekövetkezett.

Szorzási szabály több, pl. öt eseményre:

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) =$$

$$= P(A_1)\ P(A_2|A_1)\ P(A_3|A_1 \cap A_2)\ P(A_4|A_1 \cap A_2 \cap A_3)\ P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4)$$

Tehát annak a valószínűségét, hogy öt esemény mindegyike bekövetkezik, úgy lehet kiszámolni, hogy

- vesszük az első esemény valószínűségét,
- ezt megszorozzuk annak a valószínűségével, hogy a második bekövetkezik, feltéve, hogy az első bekövetkezett,
- ezután szorzunk annak a valószínűségével, hogy a harmadik bekövetkezik, feltéve, hogy az első és a második bekövetkezett,
- majd szorzunk annak a valószínűségével, hogy a negyedik bekövetkezik, feltéve, hogy az első, a második és a harmadik bekövetkezett,
- és végül szorzunk annak a valószínűségével, hogy az ötödik bekövetkezik, feltéve, hogy az első, a második, a harmadik és a negyedik bekövetkezett.

Megjegyzés: A szorzási szabályokban az események sorrendje tetszőleges lehet. Például az A, B sorrend helyett lehet a sorrend B, A:

$$P(A \cap B) = P(B \cap A) = P(B)\ P(A|B)$$

Az A, B, C sorrend helyett lehet a sorrend A, C, B:

$$P(A \cap B \cap C) = P(A \cap C \cap B) = P(A)\ P(C|A)\ P(B|A \cap C)$$

De az A, B, C sorrend helyett lehet a sorrend akár C, B, A is:

$$P(A \cap B \cap C) = P(C \cap B \cap A) = P(C)\ P(B|C)\ P(A|B \cap C)$$

Feladat: Születésnapok paradoxona. Tegyük fel, hogy n embert véletlenszerűen kiválasztunk, és belőlük alkotunk társaságot. Ezek után a társaság tagjai egymás után hangosan bemondják, hogy az év melyik napján van a születésnapjuk. (Az egyszerűség kedvéért a szökőévektől eltekintünk, az éveket 365 naposnak vesszük.) Ha n értéke nagyobb, mint 365, akkor biztos, hogy a társaság tagjai között vannak olyanok, akiknek a születésnapja egybe esik. Ha n értéke kisebb vagy egyenlő, mint 365, akkor az is lehet, hogy a születésnapok különböző napokra esnek, de az is lehet, hogy vannak olyanok, akiknek a születésnapja egybe esik.

Első a kérdés: Mi a valószínűsége annak, hogy a társaság tagjai között vannak, akiknek a születésnapja egybe esik?

Második kérdés: Mi az a legkisebb n érték, hogy ez a valószínűség $\frac{1}{2}$ -nél már nagyobb?

Megoldás: Az A_k eseményt így definiáljuk:

$$A_k = \text{a bemondáskor az első } k \text{ ember születésnapja között nincs egybeesés} \qquad (k = 1, 2, 3, \ldots)$$

Nyilvánvaló, hogy $\mathrm{P}(A_1) = 1$.

Az $A_1, A_2, A_3, \ldots$ események szűkülő sorozatot alkotnak. Abból a célból, hogy az $\mathrm{P}(A_k|A_{k-1})$ feltételes valószínűséget meghatározzuk, tegyük fel, hogy az A_{k-1} esemény bekövetkezik, vagyis az első $k-1$ bemondott születésnap különbözik. Nyilvánvaló, hogy ilyen feltétel mellett az A_k esemény akkor és csak akkor következik be, ha a k -ik születésnap különbözik az korábbi $k-1$ születésnaptól, azaz a k -ik születésnap a maradék $365-(k-1)$ nap valamelyikére esik. Ezért

$$\mathrm{P}(A_k|A_{k-1}) = \frac{365-(k-1)}{365} \qquad (k \geq 1)$$

azaz

$$\mathrm{P}(A_2|A_1) = \frac{364}{365} = 0,9973$$

$$\mathrm{P}(A_3|A_2) = \frac{363}{365} = 0,9945$$

$$\mathrm{P}(A_4|A_3) = \frac{362}{365} = 0,9918$$

$$\vdots$$

A szorzási szabállyal kapjuk, hogy

$$\begin{array}{lllllllll}
\mathrm{P}(A_1) & = & 1 & & & & & & \\
\mathrm{P}(A_2) & = & \mathrm{P}(A_1) & \mathrm{P}(A_2|A_1) & = & 1 & 0,9973 & = & 0,9973 \\
\mathrm{P}(A_3) & = & \mathrm{P}(A_2) & \mathrm{P}(A_3|A_2) & = & 0,9973 & 0,9945 & = & 0,9918 \\
\mathrm{P}(A_4) & = & \mathrm{P}(A_3) & \mathrm{P}(A_4|A_3) & = & 0,9918 & 0,9918 & = & 0,9836 \\
\vdots & & & & & & & &
\end{array}$$

Az A_k esemény komplementere:

$$\overline{A_k} = \text{a bemondáskor az első } k \text{ ember születésnapjai között van egybeesés} \qquad (k = 1, 2, 3, \ldots)$$

A komplementer események valószínűségei:

$$\begin{aligned}
&\mathrm{P}\left(\,\overline{A_1}\,\right) = 1 - \mathrm{P}(A_1) = 1 - 1 && = 0 \\
&\mathrm{P}\left(\,\overline{A_2}\,\right) = 1 - \mathrm{P}(A_2) = 1 - 0,9973 && = 0,0027 \\
&\mathrm{P}\left(\,\overline{A_3}\,\right) = 1 - \mathrm{P}(A_3) = 1 - 0,9918 && = 0,0082 \\
&\mathrm{P}\left(\,\overline{A_4}\,\right) = 1 - \mathrm{P}(A_4) = 1 - 0,9836 && = 0,0164 \\
&\vdots
\end{aligned}$$

A képletek alapján a valószínűségek numerikus értékét kiszámoljuk, és az eredményeket táblázatba rendezzük:

n	$\mathrm{P}\left(\,\overline{A_n}\,\right)$
1	0.0000
2	0.0027
3	0.0082
4	0.0164
5	0.0271
6	0.0405
7	0.0562
8	0.0743
9	0.0946
10	0.1169
$\vdots$	$\vdots$
20	0.4114
21	0.4437
22	0.4757
23	0.5073
24	0.5383
25	0.5687
26	0.5982
27	0.6269
28	0.6545
29	0.6810
30	0.7063
$\vdots$	$\vdots$
40	0.8912
41	0.9032
42	0.9140
43	0.9239
44	0.9329
45	0.9410
46	0.9483
47	0.9548
48	0.9606
49	0.9658
50	0.9704
$\vdots$	$\vdots$

A táblázatból látjuk, hogy

$$\mathrm{P}\left(\,\overline{A_{22}}\,\right) = 0,4757$$

$$\mathrm{P}\left(\overline{A_{23}} \right) = 0,5073$$

vagyis a második kérdésre a válasz: 23.

5.3. Fa-gráf valószínűségekkel súlyozva

Feltételes valószínűségekkel kapcsolatos problémák esetén gyakran segít a gondolkodásban, számolásban, ha rajzolunk a problémához kapcsolódóan egy valószínűségekkel súlyozott fa-gráfot. Egy *valószínűségekkel súlyozott fa-gráf* valami ehhez hasonlót jelent:

- Rajzolj egy papír aljára egy pontot. Ezt a pontot *a fa gyökeré*nek nevezzük.
- Most húzzál a pontból valahány – mondjuk – három szakaszt felfelé. Ezeket a szakaszokat a *gyökérből kiinduló ágak*nak nevezzük. Írd az első ág mellé – mondjuk – a 0.2 értéket, a második ág mellé – mondjuk – a 0.3 értéket, a harmadik ág mellé – mondjuk – a 0.5 értéket. Ezeket az értékeket *súlyok*nak nevezzük. Képzeld el, hogy egy bogár a gyökérből indulva mászik felfelé, és amikor egy elágazáshoz ér, a súlyoknak megfelelő valószínűségekkel választ az egyes ágak közül, hogy aztán azob az ágon másszon tovább felfelé.
- Az első ág végénél legyen megint egy elágazás – mondjuk – két ággal. Írd az ágak mellé – mondjuk – a 0.4, illetve a 0.6 súlyokat. Ilyen valószínűségekkel megy rá a bogár ezekre az ágakra.
- A második ág végénél legyen szintén egy elágazás – mondjuk – négy ággal. Írd minden ág mellé – mondjuk – a 0.25 súlyokat. Ilyen valószínűséggel megy a bogár mindenegyes ágra.
- A harmadik ág végénél is legyen egy elágazás – mondjuk – hat ággal. Írd az ágak mellé a 0.1 ; 0.1 ; 0.1 ; 0.1 ; 0.1 ; 0.5 súlyokat. Ilyen valószínűségekkel megy a bogár ezekre az ágakra.
- Vegyük észre. hogy minden elágazásnál az odaírt súlyok összege 1.
- Elképzelhetünk még további elágazásokat és súlyokat. A bogár mindig a fa teteje felé megy (távolodik a gyökértől), és minden elágazásnál az ott megadott súlyoknak megfelelő valószínűségek szerint választja meg, hogy merre menjen tovább, amíg egyszercsak nem tud továbbmenni, mert ág már nincs tovább.
- A szorzási szabály alapján kézenfekvő, hogy milyen valószínűséggel jut el a bogár a fa egyes végződéseibe: *azokat a súlyokat kell összeszorozni, melyek a gyökérből kiindulva a szóbanforgó végződéshez vezető út mentén találhatóak.*

Ha netán ez így leírva nem volt érthető vagy meggyőző, akkor – sebaj – majd órán elmondjuk, megmutatjuk, akkor szép lesz.

5.4. További szorzási szabályok

Megjegyzés: Van, amikor eseményeknek olyan sorozatával van dolgunk, hogy a sorozat bármely eleme maga után vonja a korábbiakat, azaz

$$A_1 \supset A_2 \supset A_3 \supset A_4 \supset A_5 \ldots$$

Ilyenkor azt mondjuk, hogy eseményeknek egy **csökkenő sorozat**ával van dolgunk. Egy csökkenő sorozat esetén nyilván fennállnak az alábbiak:

$$A_2 = A_1 \cap A_2$$

$$A_3 = A_1 \cap A_2 \cap A_3$$

$$A_4 = A_1 \cap A_2 \cap A_3 \cap A_4$$
$$A_5 = A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5$$

Ezért a szorzási szabály pl. öt eseményre így egyszerűsödik:

$$\mathrm{P}(A_5) = \mathrm{P}(A_1)\ \mathrm{P}(A_2|A_1)\ \mathrm{P}(A_3|A_2)\ \mathrm{P}(A_4|A_3)\ \mathrm{P}(A_5|A_4)$$

Tehát csökkenő esemény sorozat esetén annak a valószínűségét, hogy ötödik esemény bekövetkezik, úgy lehet kiszámolni, hogy

- vesszük az első esemény valószínűségét,
- ezt megszorozzuk annak a valószínűségével, hogy a második bekövetkezik, feltéve, hogy az első bekövetkezett,
- ezután szorzunk annak a valószínűségével, hogy a harmadik bekövetkezik, feltéve, hogy a második bekövetkezett,
- majd szorzunk annak a valószínűségével, hogy a negyedik bekövetkezik, feltéve, hogy a harmadik bekövetkezett,
- és végül szorzunk annak a valószínűségével, hogy az ötödik bekövetkezik, feltéve, hogy a negyedik bekövetkezett.

Megjegyzés. *(Extra tananyag)*:
Végtelen sok esemény csökkenő sorozata esetén a szorzási szabály így fest: Ha A -val jelöljük a végtelen sok esemény metszetét, akkor

$$\mathrm{P}(A) = \mathrm{P}(A_1)\ \mathrm{P}(A_2|A_1)\ \mathrm{P}(A_3|A_2) \dots$$

Tehát annak az eseménynek a valószínűségét hogy

a végtelen sok $A_1, A_2, A_3, \dots$ esemény mindegyike bekövetkezik

úgy lehet kiszámolni, hogy

- vesszük az első esemény valószínűségét,
- szorzunk a másodiknak az elsőre vonatkozó feltételes valószínűségével,
- szorzunk a harmadiknak a másodikra vonatkozó feltételes valószínűségével,
- és így tovább folytatjuk a szorozgatást a végtelenségig úgy, hogy
- mindig a soron következő eseménynek az őt megelőzőre vonatkozó feltételes valószínűségével szorzunk.

Technikailag egy ilyen végtelen sok tényezőből álló szorzatot ahhoz hasonlóan kell kezelni, mint ahogy a végtelen sok tagból álló összegeket kezeljük: a véges sok tényezőből álló

$$\mathrm{P}(A_1)\ \mathrm{P}(A_2|A_1)\ \dots\ \mathrm{P}(A_n|A_{n-1})$$

szorzatnak $n \to \infty$ mellett a határértéket veszünk, vagyis

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 \cap \dots) = \lim_{n\to\infty} \mathrm{P}(A_1)\ \mathrm{P}(A_2|A_1)\ \dots\ \mathrm{P}(A_n|A_{n-1})$$

5.5. Teljes valószínűség formulája és Bayes formula

Teljes esemény rendszer: Azt mondjuk, hogy az $A_1, A_2, \ldots, A_n$ események teljes eseményrendszert alkotnak, ha egymást kizáróak és uniójuk a biztos esemény. A teljes eseményrendszer tagjai közül a 0 valószínűségűeket eldobhatjuk, ezért felételezzük, hogy $\mathrm{P}(A_i) \neq 0$ minden i -re.

Teljes valószínűség formulája: Legyen $A_1, A_2, \ldots, A_n$ teljes eseményrendszer, B pedig tetszőleges esemény. Ekkor:

$$\mathrm{P}(B) = \mathrm{P}(A_1) \cdot \mathrm{P}(B|A_1) + \mathrm{P}(A_2) \cdot \mathrm{P}(B|A_2) + \ldots \mathrm{P}(A_n) \cdot \mathrm{P}(B|A_n)$$

Bayes formula: Ha a jelenség lezajlása során valahogyan megtudjuk, hogy egy B esemény bekövetkezett, akkor kérdezhetjük: Ez a feltétel hogyan módosítja az egyes A_i események esélyeit? A választ a $\mathrm{P}(A_i|B)$ feltételes valószínűség adja, amit az alábbi képlettel számolhatunk ki:

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(A_i \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A_i) \cdot \mathrm{P}(B|A_i)}{\mathrm{P}(A_1) \cdot \mathrm{P}(B|A_1) + \mathrm{P}(A_2) \cdot \mathrm{P}(B|A_2) + \ldots \mathrm{P}(A_n) \cdot \mathrm{P}(B|A_n)}$$

Példa: Valószínűségek helyett százalékok. Ebben a példában semmi véletlen sincs, mégis segíthet a fenti két formula megemésztésében. Tegyük fel, hogy egy ládában sok, mondjuk 1000, fából és vasból készült színes golyó van. Tegyük fel, hogy a golyók

50%-a piros,	30%-a zöld,	20%-a kék,	továbbá, hogy
a pirosak 40%-a,	a zöldek 70%-a,	a kékek 90%-a	fából készült (a többi vasból).

1. Segítség a teljes valószínűség formulájának megemésztéséhez: Könnyű kiszámolni, hogy az összes golyó hányad része készült fából:

$$0.5 * 0.4 + 0.3 * 0.7 + 0.2 * 0.9 = 0.59$$

2. Segítség a Bayes formula megemésztéséhez: Azt is könnyű látni, hogy a fából készül golyóknak

- hányad része piros:
$$\frac{0.5 * 0.4}{0.59} = \frac{0.5 * 0.4}{0.5 * 0.4 + 0.3 * 0.7 + 0.2 * 0.9} = 0.34$$
- hányad része zöld:
$$\frac{0.3 * 0.7}{0.59} = \frac{0.3 * 0.7}{0.5 * 0.4 + 0.3 * 0.7 + 0.2 * 0.9} = 0.36$$
- hányad része kék:
$$\frac{0.2 * 0.9}{0.59} = \frac{0.2 * 0.9}{0.5 * 0.4 + 0.3 * 0.7 + 0.2 * 0.9} = 0.31$$

Feladat: Dobókocka, dobozok, színes golyók. Van három dobozunk. Az elsőben egy piros és egy fehér golyó van, a másodikban két piros és egy fehér, a harmadikban egy piros és három fehér. A véletlenre bízzuk, hogy melyik dobozból húzunk egy golyót. Ha a dobókockánkkal 1-est, 2-est vagy 3-ast dobunk, akkor a első dobozból, ha 4-est vagy 5-öst, akkor másodikból, ha 6-ost, akkor a harmadikból. Kérdés: mi a valószínűsége annak, hogy a véletlenszerűen választott dobozból kihúzott golyó piros?

Megoldás: Az alábbi valószínűségek és feltételes valószínűségek a feladat szövege alapján nyilvánvalóak:

$$\mathrm{P}(\text{ első doboz }) = \frac{3}{6} \qquad \mathrm{P}(\text{ második doboz }) = \frac{2}{6} \qquad \mathrm{P}(\text{ harmadik doboz }) = \frac{1}{6}$$

$$\text{P(piros | első doboz)} = \frac{1}{2} \qquad \text{P(piros | második doboz)} = \frac{2}{3} \qquad \text{P(piros | harmadik doboz)} = \frac{1}{4}$$

A teljes valószínűség formuláját alkalmazva:

$$\begin{gathered}
\text{P(piros)} = \\
= \ \text{P(első doboz)} \ \cdot \ \text{P(piros | első doboz)} \ + \\
+ \ \text{P(második doboz)} \ \cdot \ \text{P(piros | második doboz)} \ + \\
+ \ \text{P(harmadik doboz)} \ \cdot \ \text{P(piros | harmadik doboz)} \ = \\
= \ \frac{3}{6} \cdot \frac{1}{2} \ + \ \frac{2}{6} \cdot \frac{2}{3} \ + \ \frac{1}{6} \cdot \frac{1}{4} \ = 0.514
\end{gathered}$$

Megjegyzés: A képletekkel leírt gondolatmenetet le is lehet rajzolni egy valószínűségekkel súlyozott fa-gráf segítségével. Olvasás közben tessék a rajzolni!

A fa gyökeréből kiindul három ág.

- Az első ág annak felel meg, hogy a dobókockával 1-est, 2-est vagy 3-ast dobunk, azaz az első dobozhoz jutunk.
- A második ág annak felel meg, hogy a dobókockával 4-est vagy 5-öst dobunk, azaz a második dobozhoz jutunk.
- A harmadik ág annak felel meg, hogy a dobókockával 6-ost dobunk, azaz a harmadik dobozhoz jutunk.

Az egyes ágakhoz súlyok tartoznak.

- Az első doboznak megfelelő ágon a súly $\frac{3}{6}$.
- A második doboznak megfelelő ágon a súly $\frac{2}{6}$.
- A harmadik doboznak megfelelő ágon a súly $\frac{1}{6}$.

Mindegyik ág kétfelé ágazik. Mindegyik elágazásnál az első ág a piros, a második ág a fehér húzásának felel meg.

- Az első doboznak megfelelő ág folytatásainál a súlyok: $\frac{1}{2}$ illetve $\frac{1}{2}$.
- A második doboznak megfelelő ág folytatásainál a súlyok: $\frac{2}{3}$ illetve $\frac{1}{3}$.
- A harmadik doboznak megfelelő ág folytatásainál a súlyok: $\frac{1}{4}$ illetve $\frac{3}{4}$.

Feladat *(az előző feladat folytatása:)* Tegyük fel, hogy valaki az előző feladatban leírtaknak megfelelően a másik szobában végrehajtja a kísérletet úgy, ahogy kell, aztán átjön a mi szobánkba, és közli, hogy piros golyót húzott. De nem mondja meg, hogy melyik dobozból húzta. Természetes módon merül fel bennünk a három kérdés:

1. Mi a valószínűsége annak, hogy a piros golyót az 1. dobozból húzta?
2. Mi a valószínűsége annak, hogy a piros golyót a 2. dobozból húzta?
3. Mi a valószínűsége annak, hogy a piros golyót a 3. dobozból húzta?

Megoldás:

1.
$$\text{P(1. doboz | piros)} = \frac{\text{P(1. doboz ÉS piros)}}{\text{P(piros)}} = \frac{\frac{3}{6} \cdot \frac{1}{2}}{\frac{3}{6} \cdot \frac{1}{2} + \frac{2}{6} \cdot \frac{2}{3} + \frac{1}{6} \cdot \frac{1}{4}} = 0.49$$

2.

$$\text{P(2. doboz } \mid \text{piros)} = \frac{\text{P(2. doboz ÉS piros)}}{\text{P(piros)}} = \frac{\frac{2}{6} \cdot \frac{2}{3}}{\frac{3}{6} \cdot \frac{1}{2} + \frac{2}{6} \cdot \frac{2}{3} + \frac{1}{6} \cdot \frac{1}{4}} = 0.43$$

3.

$$\text{P(3. doboz } \mid \text{piros)} = \frac{\text{P(3. doboz ÉS piros)}}{\text{P(piros)}} = \frac{\frac{1}{6} \cdot \frac{1}{4}}{\frac{3}{6} \cdot \frac{1}{2} + \frac{2}{6} \cdot \frac{2}{3} + \frac{1}{6} \cdot \frac{1}{4}} = 0.08$$

Látjuk, hogy ezek a

$$\text{P(1. doboz } \mid \text{piros)} = 0.49 \qquad \text{P(2. doboz } \mid \text{piros)} = 0.43 \qquad \text{P(3. doboz } \mid \text{piros)} = 0.08$$

feltételes valószínűségek különböznek a feltétel nélküli

$$\text{P(1. doboz)} = 0.50 \qquad \text{P(2. doboz)} = 0.33 \qquad \text{P(3. doboz)} = 0.17$$

valószínűségektől.

Feladat *(az előző feladat folytatása:)* És mi van akkor, ha az ember azt közli, hogy fehér golyót húzott? A három kérdés most:

- Mi a valószínűsége annak, hogy a fehér golyót az 1. dobozból húzta?
- Mi a valószínűsége annak, hogy a fehér golyót a 2. dobozból húzta?
- Mi a valószínűsége annak, hogy a fehér golyót a 3. dobozból húzta?

Megoldás: Az Olvasó könnyen ellenőrizheti, hogy

$$\text{P(1. doboz } \mid \text{fehér)} = 0.51 \qquad \text{P(2. doboz } \mid \text{fehér)} = 0.23 \qquad \text{P(3. doboz } \mid \text{fehér)} = 0.26$$

Ezek a valószínűség értékek is különböznek azoktól, melyeket az előző megoldás végén kiírtunk.

5.6. Optimális taktika előre nem látható helyzetekben *(Extra tananyag)*

Számtalanszor fordul elő az életünkben, hogy előre nem látható helyzetekben kell döntenünk. Például amikor használt autót akarunk venni (természetesen jót és olcsót!), és meglehetősen tapasztalatlanul kezdjük nézegetni a kínálatot, akkor egy idegig csak nézegetünk, "szimatolgatunk", aztán egyszercsak ráakadunk egy olyan vételi lehetőségre, ami jobbnak tűnik, mint az összes korábbi, és akkor erre gyorsan lecsapunk, és megvesszük.

Egy ilyen probléma természtesen nagyon összetett lehet: a véletlen jelentős szerepet játszhat benne, és pszichológiai és sok egyéb tényező is beleszólhat a problémába. Mégis, egy egyszerű valószínűségszámítási modell segítségével meglepően érdekes és szép eredményre juthatunk. Ezzel a modellel foglalkozunk ebben a részben. Előkészítésként vesszük az alábbi problémát:

Tekintsünk 10 cédulát, 1-től 10-ig számozva őket. A 10-es cédulát, a legnagyobb számot nevezzük: *királynő*nek. Rakjuk le a cédulákat balról jobbra egymás mellé véletlenszerűen, azaz tekintsünk egy véletlenszerű permutációt. Például egy lehetséges permutáció:

$$6,\ 5,\ 7,\ 4,\ 1,\ 8,\ 2,\ 10,\ 9,\ 3$$

A királynő, ebben a permutációban a 8-ik pozícióra került. Keressük meg most a királynő előtti legnagyobb számot! Ez most a 8-as, ami a 6-ik pozíción áll. A királynő előtti legnagyobb számot *szolgáló lány*nak nevezzük. Tehát most a szolgáló lány a 8-as, és ő a 6-ik pozíción áll. Jelöljük X -szel a királynő pozícióját, Y -nal a szolgáló lány pozícióját.

A fenti példában tehát $X = 8$, $Y = 6$. Ha $X = 1$, azaz a királynő az 1-ső pozíción áll, akkor nincs szolgálólány, ilyenkor Y értékét 0-nak vesszük.

Segédfeladat: Rögzítsünk egy c számot, mely eleget tesz az $1 \le c \le 9$ egyenlőtlenségeknek. Kiszámoljuk a

$$\mathrm{P}(\ X > c \quad \text{és} \quad Y \le c\ )$$

valószínűséget, vagyis annak az eseménynek a valószínűségét, hogy a királynő pozíciója nagyobb c -nél, és a szolgálólány pozíciója kisebb vagy egyenlő c -nél. Ez a valószínűség – természetesen – függ c -től, ezért a valószínűséget c -vel kifejezve adjuk meg.

Megoldás:

$$\begin{aligned}
&\mathrm{P}(\ X > c \quad \text{és} \quad Y \le c\ ) \\
&= \sum_{k=c+1}^{10} \mathrm{P}(\ X = k \quad \text{és} \quad Y \le c\ ) \\
&= \sum_{k=c+1}^{10} \mathrm{P}(\ X = k\ )\ \mathrm{P}(\ Y \le c \mid X = k\ ) \\
&= \sum_{k=c+1}^{10} \frac{1}{10}\ \frac{c}{k-1} \\
&= \frac{1}{10} \sum_{k=c+1}^{10} \frac{c}{k-1}
\end{aligned}$$

Megjegyzés: 10 cédula helyett vegyünk most 100 cédulát, megszámozva őket 1-től 100-ig. Most a 100 -as cédulát hívjuk királynőnek. Ha lerakjuk a 100 cédulát balról jobbra egymás mellé véletlenszerűen, azaz tekintünk egy véletlenszerű permutációt, akkor a királynő előtti számok között a legnagyobbat hívjuk szolgálólánynak. Jelöljük X -szel a királynő pozícióját, Y -nal a szolgáló lány pozícióját. Ha $X = 1$, azaz a királynő az 1-ső pozíción áll, akkor nincs szolgálólány, ilyenkor Y értékét 0-nak vesszük. Válasszunk most is egy c számot, mely eleget tesz az $1 \le c \le 99$ egyenlőtlenségeknek.

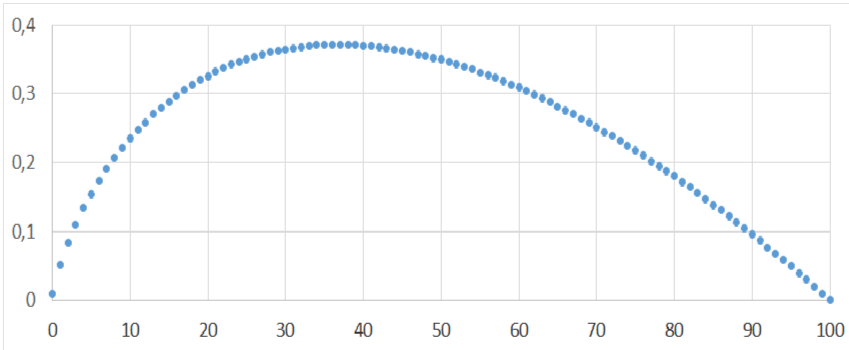
100 cédula esetén is kiszámoljuk a a hasonlóképpen értelmezett

$$\mathrm{P}(\ X > c \quad \text{és} \quad Y \le c\ )$$

valószínűséget. Az eredmény kézenfekvő:

$$\mathrm{P}(\ X > c \quad \text{és} \quad Y \le c\ ) = \frac{1}{100} \sum_{k=c+1}^{100} \frac{c}{k-1}$$

Ez a képlet fontos lesz a következő példa megoldásában, ezért minden 0 és 100 közötti c -re kiszámoltuk Excellel, és a numerikus eredményeket táblázatba foglatuk. A táblázat alapján grafikont is készítettünk, ezt láthatjuk "A kiszámított valószínűség c függvényében" című ábrán. Magát a táblázatot – a hellyel való spórolás miatt – csak rövidítve adjuk meg. A táblázatnak csupán az elejét, a közepének egy részét és a végét mutatjuk:



1. ábra. *A kiszámított valószínűség c függvényében*

c	$\frac{1}{100} \sum_{k=c+1}^{100} \frac{c}{k-1}$
0	0.010
1	0.052
2	0.084
3	0.110
4	0.134
5	0.155
$\vdots$	$\vdots$
35	0.37071
36	0.37101
37	0.37104
38	0.37080
39	0.37030
40	0.36953
$\vdots$	$\vdots$
95	0.049
96	0.039
97	0.030
98	0.020
99	0.010
100	0.000

Az ábráról és a táblázatból látjuk, hogy a maximális valószínűség $c = 37$ -nél adódik, és hogy a maximális valószínűség értéke hat tizedesre kerekítve 0.371043, két tizedesre kerekítve 0.37. Ezt a tényt használni fogjuk a következő – világhírnek örvendő – feladatban.

Feladat: Szindbad és a háremhölgyek. Egyszer a török szultán – jutalomképpen – felajánlotta Szindbádnak, a híres nőcsábásznak, hogy 100 gyönyörű hármehölgye közül választhat egyet, és a kiválsztott hölggyel eltölhet egy éjszakát.

Szindbád soha nem látta korábban a hölgyeket, és most is csak nagyon korlátozott körülmények között találkozhat velük: a hölgyek egyesével jelennek meg Szindbád előtt véletlenszerű sorrendben. Minden lehetséges sorrend ugyanakkora valószínűségű – ezt a tényt Szindbádnak előre megmondták. Szindbád csak egyszer mondhatja az előtte éppen mutatkozó hölgyre, hogy "őt választom". Akit Szindbád nem választ ki a megjelenésekor, azt már később "visszamenőleg" nem kérheti. Választását később nem módosíthaja.

A háremhölgyek között létezik egy jól definiált – mindenki számára nyilvánvaló – szépség sorrend: van közöttük egy legszebb, egy második legszebb, és így tovább egy 100-ik legszebb. Bármely két hölgy esetében egyértelmű, hogy melyikük a szebb. Szindbád, aki – ismételjük – nem ismeri a hölgyeket, és most is csak a megjelenésükkor látja őket, arra törekszik, hogy elcsípje a legszebbet. Ha a procedúra végén kiderül, hogy ez sikerült neki, akkor – hiúsága beteljesül, és akcióját sikeresnek könyveli el, ha nem, akkor az akció nem ért semmit a számára . (Szindbád ilyen. Úgy kell neki!)

Hogyan válasszon Szindbád? Úgy tűnhet, hogy a siker elérésének nagyon kicsi az esélye. Valóban, ha csak úgy véletlenszerűen választ egy hölgyet, mondjuk a legelsőt, vagy kisorsolja előre, hogy hányadikat, akkor 0.01 az esélye annak, hogy a legszebb jut neki.

Ha viszont okosan taktikázik, akkor a bölcsesség meglepően nagy valószínűséggel sikerre viheti az akcióját! Ez fog kiderülni az itt következő megoldásból! Ilyen bölcsességek jól jönnek mindenkinek – ezért nézzük máris a megoldást!

Megoldás: Szindbad így gondolkodhat: Képzeletben választ egy c számot $(0 \le c \le 99)$, és előre eldöni magában, hogy az első c hölgy közül semmiképpen sem választ, csak megfigyeli őket, és megjegyzi, milyen szép volt közülük a legszebb. Aztán a $(c+1)$ -iknek megjelenő hölgytől kezdve már csak arra figyel, hogy felbukkan-e olyan hölgy, akinek szépsége meghaladja az első c hölgy során kifigyelt maximumális szépséget. Ha felbukkan ilyen hölgy, akkor arra lecsap, ha nem, akkor nem választ senkit. Nevezzük ezt a taktikát c -*taktiká*nak!

A segédfeladatunkra támaszkodva könnyű megadni, hogy ezzel a c -taktikával mi a valószínűsége annak, hogy Szindbád elcsípi a legszebb hölgyet. Felhasználva a korábban definiált királynő és szolgálólány fogalmát, és a velük kapcsolatban bevezetett X és Y valószínűségi változókat, azt kell az Olvasónak meggondolni, hogy

az az esemény, hogy

Szindbád a c -taktikával elcsípi a legszebb hölgyet

akkor és csak akkor következik be, ha

a királynő pozíciója $> c$ *és a szolgálólány pozíciója* $\le c$

esemény bekövetkezik.

Ennek az eseménynek a valószínűségét a segédfeladathoz fűzött megjegyzésben megadtuk:

$$\frac{c}{100} \sum_{k=c+1}^{100} \frac{1}{k-1}$$

A képlettel kapcsolatban készített táblázatból és ábrából világosan kiderül, hogy ha Szindbád a $c = 37$ -es taktikát alkalmazza, akkor 0.37 valószínűséggel elcsípi a legszebb hölgyet.

A 0.37 valószínűség – bizony – nincs túl közel az áhított 1-hez, de

- lényegesen nagyobb az "ész nélküli taktika" 0.01 -es siker valószínűségénél, és
- be lehet bizonyítani, hogy nem létezik olyan taktika, ami 0.37 -nél nagyobb valószínűséggel vezetne sikerre.

Be lehet látni, hogy hasonló helyzetekben is a követendő optimális taktika így fest: a lehetőségek 37 % -át csak nézegetni kell úgy, hogy közülük nem választunk, de megjegyezzük a "szépség", "jóság" maximumát, a 37 % elengedése után viszont arra kell lecsapni, amikor a "szépség", "jóság" meghaladja a korábban kifigyelt maximumot. Ezzel a taktikával 0.37 valószínűséggel elcsípjük a lehetőségek közül a "legszebbet", "legjobbat".

5.7. Feltételes eloszlás egy eseményen belül

Példa: Fiatal házaspár egy vagy több gyerekkel. Korábbi példánk szerint, ha egy fiatal házaspárt véletlenszerűen választunk, és tekintjük az

$$X = \text{ gyerekek száma}$$

valószínűségi változót, akkor a valószínűségi változó eloszlása:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

Egyszerű összeadással kapjuk, hogy a $X \geq 1$ esemény valószínűsége $P(X \geq 1) = 0.4 + 0.3 + 0.1 = 0.8$.

Ha a véletlenszerűen választott fiatal párról tudjuk, hogy van gyerekük, de nem tudjuk, hogy hány, akkor a gyerekek számának lehetséges értékei 1, 2, 3. Ezeknek a lehetséges értékeknek a feltételes valószínűségeit táblázatba rendezve jutunk az X valószínűségi változó $X \geq 1$ **feltétel melletti (feltételen belüli) feltételes eloszlás**ához:

x	1	2	3
$p(x)$	$\frac{0.4}{0.8}$	$\frac{0.3}{0.8}$	$\frac{0.1}{0.8}$

azaz

x	1	2	3
$p(x)$	$\frac{4}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Általánosan is megfogalmazzuk a példában leírtakat: Egy

$$\{p(x),\ x \in S\}$$

eloszlás és egy $A \subseteq S$ halmaz esetén a

$$\{\frac{p(x)}{P(A)},\ x \in A\}$$

eloszlást **az** A **esemény bekövekezése melletti feltételes eloszlás**ának nevezzük. Ha az eredeti eloszlás egy X valószínűségi változó eloszlása, akkor a feltételes eloszlást **az** X **valószínűségi változónak az** A **-n belüli eloszolásá**nak is hívjuk.

5.8. Feltételes eloszlások rendszere síkbeli eloszlás esetén

Először konkrét példákat veszünk, aztán megismerkedünk a feltételes eloszlások rendszerének általános fogalmával.

5.8.1. Példák

1. Példa: Hány kék, ha tudjuk, hogy hány piros? Az 1. fejezet 7. alfejezében az 5. példában feltettük, hogy egy dobozban 45 darab golyó van, melyek közül 10 darab piros, 15 darab kék, 20 darab fehér. Kiveszünk 8 darab golyót visszatevés nélkül, és megfigyeljük, hogy a kivett golyók között hány piros és hány kék van:

$$X = \text{ ahány piros van a kivett golyók között}$$

$$Y = \text{ ahány kék van a kivett golyók között}$$

Ott meghatároztuk az $(X;Y)$ kétdimenziós valószínűségi változó eloszlását, egy táblázatot kaptunk. Gondoljuk meg ennek a példának a kapcsán, hogy hogyan lehet $(X;Y)$ eloszlásából meghatározni Y feltételes eloszlását olyan feltétel mellett, hogy X értéke adott konkrét x érték?

1. Megoldás: A feltételes eloszlások tagjai nyilván feltételes valószínűségek, melyeket hányadosként kapunk a megfelelő valószínűségekből. Az alábbi táblázat minden oszlopának minden elemét úgy kaptuk meg, hogy $(X;Y)$ eloszlásának táblázatából a megfelelő $\mathrm{P}(X=x, Y=y)$ valószínűséget elosztottuk X eloszlásának táblázatából a megfelelő $\mathrm{P}(X=x)$ tagjával:

Y feltételes eloszlásai adott X értékek mellett:

y										
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
7	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
6	0.040	0.015	0.003	0.000	0.000	0.000	0.000	0.000	0.000	
5	0.145	0.085	0.037	0.009	0.000	0.000	0.000	0.000	0.000	
4	0.281	0.231	0.160	0.084	0.026	0.000	0.000	0.000	0.000	
3	0.300	0.328	0.320	0.266	0.174	0.070	0.000	0.000	0.000	
2	0.173	0.242	0.313	0.369	0.381	0.321	0.176	0.000	0.000	
1	0.049	0.086	0.143	0.224	0.327	0.435	0.504	0.429	0.000	
0	0.005	0.012	0.024	0.048	0.093	0.174	0.319	0.571	1.000	
	0	1	2	3	4	5	6	7	8	x

Megjegyzés: Ebben a táblázatban – természetesen – minden opszlopra igaz, hogy az oszlopban lévő számok összege 1 -gyel egyenlő.

2. Megoldás: Ha 45 darab golyó van egy dobozban, közülük 10 darab piros, 15 darab kék, 20 darab fehér, kiveszünk 8 darab golyót visszatevés nélkül, és kiderül, hogy a kivett golyók között 3 piros van, azaz teljesül az $X=3$ feltétel, akkor eme feltétel mellett a kék golyók számánáka az eloszlása nyilván olyan, mint ha 35 darab golyó lenne a dobozban, közülük 15 darab lenne kék, 20 darab lenne fehér, és kivennénk 5 darab golyót visszatevés nélkül. Ezért a mondott feltétel mellett annak a valószínűsége, hogy – mondjuk – 2 kék golyó lesz a kivett golyók között, vagyis annak a valószínűsége, hogy $Y=2$, a jól ismert képlet:

$$\frac{\binom{15}{2}\binom{20}{5-2}}{\binom{35}{5}}$$

Hasonlóan, annak a valószínűsége, hogy y darab kék golyó lesz a kivett golyók között, vagyis annak a valószínűsége, hogy $Y = y$, ugyanaz a képlet, csak a 2-es helyére y -t írunk:

$$\frac{\binom{15}{y}\binom{20}{5-y}}{\binom{35}{5}}$$

Ha az $X = 3$ feltételt kicseréljük az általánosabb $X = x$ feltételre, akkor a keresett feltételes valószínűséget így adhajuk meg:

$$\frac{\binom{15}{y}\binom{20}{8-x-y}}{\binom{35}{8-x}}$$

2. Példa: Hány piros, ha tudjuk, hogy hány kék? Gondoljuk meg, hogy hogyan lehet $(X;Y)$ eloszlásából meghatározni X feltételes eloszlását olyan feltétel mellett, hogy Y értéke adott konkrét y érték.

1. Megoldás: A feltételes eloszlások tagjai nyilván feltételes valószínűségek, melyeket hányadosként kapunk a megfelelő valószínűségekből. Az alábbi táblázat minden oszlopának minden elemét úgy kaptuk meg, hogy $(X;Y)$ eloszlásának táblázatából a megfelelő P$(X = x, Y = y)$ valószínűséget elosztottuk Y eloszlásának táblázatából a megfelelő P$(Y = y)$ tagjával:

X feltételes eloszlásai adott Y értékek mellett:

y										
8	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
7	0.667	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
6	0.437	0.460	0.103	0.000	0.000	0.000	0.000	0.000	0.000	
5	0.281	0.468	0.222	0.030	0.000	0.000	0.000	0.000	0.000	
4	0.177	0.416	0.312	0.088	0.008	0.000	0.000	0.000	0.000	
3	0.109	0.340	0.360	0.160	0.029	0.002	0.000	0.000	0.000	
2	0.065	0.261	0.367	0.230	0.067	0.008	0.000	0.000	0.000	
1	0.038	0.190	0.343	0.286	0.118	0.024	0.002	0.000	0.000	
0	0.022	0.132	0.298	0.318	0.174	0.049	0.007	0.000	0.000	
	0	1	2	3	4	5	6	7	8	x

Megjegyzés: Ebben a táblázatban – természetesen – minden sorra igaz, hogy a soran lévő számok összege 1.

Az előző feladat második megoldásához hasonlóan kapjuk, hogy az $Y = y$ feltétel mellett az $X = x$ esemény feltételes valószínűsége:

$$\frac{\binom{10}{x}\binom{20}{8-y-x}}{\binom{30}{8-y}}$$

5.8.2. Általános összefüggések

Ha X és Y diszkrét valószínűségi változók, akkor együttesen egy kétdimenziós diszkrét (X,Y) valószínűségi változót határoznak meg. Ennek a kétdimenziós (X,Y) valószínűségi változónak a súlyfüggvényét jelöljük $p(x,y)$ -nal, vagyis

$$p(x,y) = \mathrm{P}\left(X = x, Y = y\right)$$

X súlyfüggvényét $p_1(x)$ -szel, Y súlyfüggvényét $p_2(y)$ -nal jelöljük.

Feltételes súlyfüggvény. Ha feltételezzük, hogy $X = x$, akkor az Y valószínűségi változó $p_{2|1}(y|x)$ feltételes súlyfüggvénye osztással adódik:

$$p_{2|1}(y|x) = \frac{p(x,y)}{p_1(x)}$$

Hasonlóan, ha feltételezzük, hogy $Y = y$, akkor az X valószínűségi változó $p_{1|2}(x|y)$ feltételes súlyfüggvénye is osztással adódik:

$$p_{1|2}(x|y) = \frac{p(x,y)}{p_2(y)}$$

Annak a valószínűsége, hogy Y értéke egy $[y_1, y_2]$ intervallumba esik, feltéve, hogy $X = x$, nyilván így számolható ki:

$$\mathrm{P}(\ y_1 \le Y \le y_2 \mid X = x\ ) = \sum_{y=y_1}^{y_2} p_{2|1}(y|x)$$

Hasonlóan, annak a valószínűsége, hogy X értéke egy $[x_1, x_2]$ intervallumba esik, feltéve, hogy $Y = y$, nyilván így számolható ki:

$$\mathrm{P}(\ x_1 \le X \le x_2 \mid Y = y\ ) = \sum_{x=x_1}^{x_2} p_{1|2}(x|y)$$

Feltételes eloszlások (súlyfüggvények) rendszere.

Ha minden x mellett tekintjük a $p_{2|1}(y|x)$ feltételes eloszlást (súlyfüggvényt), akkor eloszlásoknak (súlyfüggvényeknek)egy rendszerét kapjuk, amit az Y valószínűségi változó X -re vonatkozó **feltételes eloszlásai (súlyfüggvényei) rendszerének** nevezünk.

Ha minden y mellett tekintjük a $p_{1|2}(x|y)$ feltételes eloszlást (súlyfüggvényt), akkor eloszlásoknak (súlyfüggvényeknek)egy rendszerét kapjuk, amit az X valószínűségi változó Y -re vonatkozó **feltételes eloszlásai (súlyfüggvényei) rendszerének** nevezünk.

5.9. Óvakodjunk a félreérthető feladatoktól!

A valószínűségszámítás tanulása során szinte biztos, hogy a tanuló találkozik olyan feladattal, melyben a feltételek, körülmények nincsenek pontosan megadva. Ebből persze félreértések adódnak, viták támadnak, melyek során a diák ömbizalma alaposan sérülhet. Erre a negatív lehetőségre mutatunk itt egy példát.

Feladat: Ha van fiú, van-e lány is? *(Ilyen címmel szerepelt már egy feladat ennek a fejezetnek az elején, az 5.1 alfejezetben. Most ennek a feladatnak egy módosításával egy másik, de a korábbihoz hasonló, eléggé elterjedt és népszerű változatával foglalkozunk. Amikor két hasonló de mégis különböző feladatot összekevernek, összemossnak, vagy egy feladat nincs pontosan megfogalmazva, és ezért többféleképpen is lehet azt érteni, komoly nézeteltérések, viták "élet-halál harcok" szoktak támadni még valószínűségszámításban járatos emberek között is!)*

Íme a maga az új feladat:*(Ajtó nyitogatós változat.)* Egy véletlenszerűen választott ismeretlen kétgyerekes családhoz becsönget valaki. A sors úgy hozza, hogy fiú nyit ajtót. Mi a valószínűsége annak, hogy a testvére lány, vagyis a fiú mellett lány is van a családban?

A) Megoldás: Ha fiú nyit ajtót, akkor ez a tény jelzi, hogy a családban van fiú. A fejezet elején kiszámoltuk, hogy ilyen feltétel mellett annak a valószínűsége, hogy lány is van a családban, $2/3$ -dal egyenlő:

$$\text{P(van lány a családban | fiú nyit ajtót)} = 2/3$$

B) Megoldás: Az ajtót nyitó gyerek testvérének neme független az ajtót nyitó gyerek nemétől. Ezért annak a valószínűsége, hogy lány is van a családban, $1/2$ -del egyenlő:

$$\text{P(van lány a családban | fiú nyit ajtót)} = 1/2$$

Kérdés: Melyik a jó megoldás?

Válasz: Mindkét megoldás jó lehet, ha a feladatot megfeleően pontosítjuk.

1. Ha olyan társadalomban élünk, ahol a fiúk udvariasak, és – fiú-lány testvérpár esetén – lánytesvérüket megelőzve pattannak ajtót nyitni, akkor az A) megoldás a jó:

   $$\text{P(van lány a családban | fiú nyit ajtót)} = 2/3$$

2. Ha mindig az idősebb gyerek megy ajtót nyitni, vagy ha mindig a fiatalabb, vagy ha igazságos sorsolással döntik el, hogy ki menjen ajtót nyitni, akkor a B) megoldás a jó:

   $$\text{P(van lány a családban | fiú nyit ajtót)} = 1/2$$

3. Ha pedig a társadalmi szokások miatt – fiú-lány testvérpár esetén – a lányok nyitnak ajtót, akkor nyilván a kérdéses feltételes valószínűség 0:

   $$\text{P(van lány a családban | fiú nyit ajtót)} = 0$$

4. Kis fantáziával – ennyi fantázia a valószínűségszámítás tanulása során nélkülözhetetlen! – el lehet képzelni olyan társadalmat is, ahol fiú-lány testvérpár esetén p valószínűséggel a fiú, $(1-p)$ valószínűséggel a lány megy ajtót nyitni.

   Ez a "véletlenítés" megvalósulhat például úgy, hogy

   (a) p értékének megfelelően a fiú és a lány sorsot húz, és úgy is, hogy
   (b) feltételezzük, hogy a társadalomban a kétgyerekes családok p -ed részében a fiú az ajtó nyitogató, $(1-p)$ -ed részében a lány.

   Ekkor – mint mindjárt megmutatjuk – a szóbanforgó feltételes valószínűség 0 és $2/3$ között akármilyen értékű is lehet, hiszen az értékét az alábbi képlet adja meg:

   $$\text{P(van lány a családban | fiú nyit ajtót)} = \frac{2p}{1+2p}$$

   Nyilvánvaló, hogy $p = 0$ esetén ez a képlet 0 -t ad, $p = 1$ esetén $2/3$ -ot.

   A képlet levezetése:

   $$\text{P(van lány | fiú nyit ajtót)} =$$

   $$= \frac{\text{P(van lány ÉS fiú nyit ajtót)}}{\text{P(fiú nyit ajtót)}}$$

   $$= \frac{\text{P(van lány is és fiú is ÉS fiú nyit ajtót)}}{\text{P(fiú nyit ajtót)}}$$

$$= \frac{\text{P(van lány is és fiú is)} \cdot \text{P(fiú nyit ajtót | van lány is és fiú is)}}{\text{P(fiú nyit ajtót)}}$$

$$= \frac{\text{P(van lány is és fiú is)} \cdot \text{P(fiú nyit ajtót | van lány is és fiú is)}}{\text{P(két fiú van)} \cdot \text{P(fiú nyit ajtót | két fiú van)} + \text{P(van lány is és fiú is)} \cdot \text{P(fiú nyit ajtót | van lány is és fiú is)}}$$

$$= \frac{0.5 \cdot p}{0.25 \cdot 1 + \ 0.5 \cdot p} = \frac{2\,p}{1\ +\ 2\,p}$$

Konklúzió: Jobb, ha elkerüljük az ilyen – nem kellően pontosan definiált – feladatokat! Természetesen az is jó – ha van rá idő és lehetőség – ha megfelelő nehézségű példákon megtanítják a diákoknak, hogy egyrészt észrevegyék, amikor egy feladatban a feltételek nincsenek pontosan megadva, másrészt érezzék, hogy mit is kell pontosítani ahhoz, a hogy a feladat korrekt legyen.

5.10. Gyakorló feladatok

I. téma: Feltételes valószínűség

1. Egy szabályos dobókockával dobunk. Barátom látja a dobás eredményét, de én nem. Mennyi a valószínűsége, hogy 6 -ost dobtunk,
   (a) ha barátomtól tudom, hogy párosat dobtunk?
   (b) És ha azt tudom, hogy legalább 3 -ast dobtunk?
   (c) És ha azt tudom, hogy legfeljebb 5 -öst dobtunk?
2. Feldobunk két dobókockát.
   (a) Mi annak a valószínűsége, hogy legalább az egyik kockán 2 -est dobunk?
   (b) Mi annak a valószínűsége, hogy legalább az egyik kockán 2 -est dobunk, ha már tudjuk, hogy a dobott számok összege 6?
3. Tegyük fel, hogy születendő gyermek azonos eséllyel lesz fiú vagy lány. Tekinsünk egy véletlenszerűen választott háromgyerekes családot. Ha megtudjuk, hogy a családban van fiú, akkor mennyi annak a valószínűsége, hogy
   (a) pontosan egy fiú van a családban?
   (b) pontosan két fiú van a családban?
   (c) pontosan három fiú van a családban?
4. A barátommal snapszerozom. Ebben a játékban 20 darab lap van, mind a négy színből öt. Én is és barátom is kap 5 lapot.
   (a) Mi a valószínűsége, hogy a barátomnak van zöldje?
   (b) Mi a valószínűsége, hogy a barátomnak van zöldje, ha nekem 3 zöldem és 2 pirosam van?

5. Egy iskolába 260 ember jár, 230 tanuló és 30 tanár. Egyszer egy influenzajárvány tört ki köztük. Az orvos az alábbi táblázatot készítette:

	beteg	egészséges	**összesen**	esemény jele
fiú	50	60	**110**	A_1
lány	40	80	**120**	A_2
tanár	10	20	**30**	A_3
összesen	**100**	**160**	**260**	
esemény jele	B_1	B_2		

(a) Véletlenszerűen kihúzunk egy kartont. Mi a valószínűsége, hogy a karton

i. fiúé?

ii. betegé?

iii. beteg fiúé?

(b) Ha előzetesen a fiúk, lányok és tanárok kartonjait külön fiókokba gyűjtötték, én a lányokéból húzok, mi a valószínűsége annak, hogy beteg lányt húztam?

(c) Az orvos szorgos asszistense egy kupacba kidobálta a fiókokból az összes kartont, aki beteg volt. Ebből véletlenszerűen húzva egyet, mi a valószínűsége annak, hogy tanár az illető?

(d) Ha kettőt húzok ugyanebből a beteg-kupacból egymás után, mi a valószínűsége, hogy az első fiú lesz, a második lány? És hogy mindkettő fiú lesz?

II. téma: Szorzási szabály

6. Egy urnában 3 piros, 5 fehér és 6 zöld golyó van. Kihúzunk közülük 3 golyót. Mennyi a valószínűsége, hogy elsőre pirosat, másodikra fehéret, harmadikra zöldet húzunk, ha húzás után a golyókat

(a) visszatesszük

(b) nem tesszük vissza?

7. Egy lakótelepen csótányirtást végeztek. Az első vegykezelés még a csótányok 60%-át irtja ki, de utána a csótányok egyre inkább immúnissá válnak, így a másodszorra már csak a 40%, harmadszorra pedig csak a 20%-uk pusztul el. Mi a valószínűsége, hogy egy megjelölt csótány

(a) átvészeli a teljes eljárást?

(b) az utolsó irtáskor pusztul el?

(c) túléli a kezelést, ha az első kezelés után még látták élve?

8. Egy dobozban 16 alkatrész közül 3 hibás. Mi a valószínűsége, hogy három egymás után kivett alkatrész működőképes?

9. Egy valszámvizsgán 30 tétel van. Ezek közül 6 a nevezetes eloszlásokkal kapcsolatos. Az első két szóbeliző hallgató kihúz egy-egy tételt. Mi annak a valószínűsége, hogy

(a) csak az első hallgató húz nevezetes eloszlásos tételt?

(b) mindkét hallgató ilyen tételt húz (húzhatják mindketten ugyanazt is!)

(c) egyik sem húz ilyen tételt?

10. Van két dobozunk. Az egyik dobozban 3 piros és 2 kék golyó van, a másikban 2 piros és 4 kék. Az első dobozból átrakunk egy golyót a másik dobozba, majd onnan átrakunk egyet az elsőbe.

(a) Mi a valószínűsége annak, hogy ezek után az első dobozban 3 piros golyó lesz?

(b) Mi a valószínűsége annak, hogy az első golyó piros és a második kék?
(c) Mi a valószínűsége annak, hogy az első golyó piros volt, ha a második kék?

11. Van két dobozunk. Az egyik dobozban 3 piros és 2 kék golyó van, a másikban 2 piros és 4 kék. Az első dobozból átrakunk egy golyót a másik dobozba, majd onnan átrakunk egyet az elsőbe, majd az első dobozból ismét egyet a másikba. Találjon ki olyan kérdéseket, melyek golyók színével kapcsolatosak, és feltételes valószínűségekkel megválaszolhatók!

III. téma: Teljes valószínűség tétele és Bayes tétel

12. Egy suli tanulóinak 80%-a lány. Az első matekvizsgán általában a lányok 85%-a, a fiúk 90%-a megy át. A hallgatóságnak kb. hány százaléka megy át az első vizsgán?

13. Két szabályos dobókockával dobunk, és megnézzük a dobott számok különbségét (ami egy 0 és 5 közötti szám). Amennyi a dobott számok különbsége, annyi szabályos érmével dobunk. (Tehát az is lehet, hogy 0 darab érmével dobunk.) Mi a valószínűsége annak, hogy az érmékkel pontosan 4 fejet kapunk?

14. Első lépés: három tízforintos érmével dobunk, és megnézzük, hány fejet kapunk. Második lépés: ahány fejet kaptunk az első lépésben, annyi húszforintos érmével dobunk. Mi a valószínűsége annak, hogy a húszforintos érmékkel pontosan 2 fejet kapunk,

(a) ha nem tudjuk, hogy a tízforintos érmékkel mi jött ki?
(b) ha tudjuk, hogy a tízforintos érmékkel legalább 2 fejet kaptunk?

15. A ketyere gyárban az A, B és C gépsoron állítják elő a ketyeréket. Az A gépsoron a ketyerék 25, a B -n 35, a C -n 40%-át gyártják. Az A gépsoron előállított ketyerék 5%-a, a B gépsoron előállítottak 4%-a, a C -n gyártott ketyeréknek csak 2%-a hibás. A hibásakat félredobják egy nagy kupacba. Ebből véletlenszerűen kiszedve egy ketyerét, mi a valószínűsége, hogy azt az A, B, illetve a C gépsoron gyártották?

16. Egy bináris csatornán a 0 jelet $1/3$, az 1 jelet $2/3$ valószínűséggel adják le. Mivel az adást zajok zavarják, ha 0 -t adnak le, akkor $1/4$ valószínűséggel 1 érkezik, ha pedig 1 -et adnak le, $1/5$ valószínűséggel 0 érkezik.

(a) Kaptunk egy 0 -t. Mi a valószínűsége, hogy ezt 0 -ként is adták le?
(b) Mi a valószínűsége, hogy 1 -et kapunk?

17. Egy dobozban pénz érmék vannak. Az érmék súlyai:

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**
tömeg (g)	4.2	6.1	7.0	7.7	8.0	9.5

18. Feltételezzük, hogy húzáskor az érmék valószínűségei arányosak a súlyaikkal. Háromszor húzunk úgy, hogy az 5, 10, 20 és 50 forintos érméket visszatesszük, a 100 és 200 forintos érméket nem. Mi a valószínűsége annak, hogy

(a) a 3 -ik húzásnál 100 vagy 200 forintos érmét húzunk?
(b) a 2 -ik húzásnál 100 vagy 200 forintos érmét húzunk, feltéve, hogy a 3 -ik húzásnál 100 vagy 200 forintos érmét húzunk?

19. Ugyanaz a helyzet mint az előző feladatban, de most – kezdetkor – minden érméből 3 -at teszünk a dobozba. Most mi a valószínűsége annak, hogy

(a) a 3 -ik húzásnál 100 vagy 200 forintos érmét húzunk?
(b) a 2 -ik húzásnál 100 vagy 200 forintos érmét húzunk, feltéve, hogy a 3 -ik húzásnál 100 vagy 200 forintos érmét húzunk?

6. Függetlenség

6.1. Események függetlensége

Két esemény függetlensége: Azt mondjuk, hogy a B, $\overline{B}$ eseménypár független az A, $\overline{A}$ eseménypártól, vagy rövidebben csak annyit mondunk, hogy a B esemény független az A eseménytől, ha az A vagy az $\overline{A}$ bekövetkezése nem módosítja a B, $\overline{B}$ eseménypár valószínűségeit, azaz

$$\begin{aligned} \mathrm{P}(\ B \mid A\ ) &= \mathrm{P}(\ B\ ) \\ \mathrm{P}(\ \overline{B} \mid A\ ) &= \mathrm{P}(\ \overline{B}\ ) \\ \mathrm{P}(\ B \mid \overline{A}\ ) &= \mathrm{P}(\ B\ ) \\ \mathrm{P}(\ \overline{B} \mid \overline{A}\ ) &= \mathrm{P}(\ \overline{B}\ ) \end{aligned}$$

B -nek A -tól való függetlensége azt fejezi ki, hogy A bekövetkezése, illetve nem bekövetkezése sem nem növeli, sem nem csökkenti a B bekövetkezésének az esélyét.

Nem nehéz megmutatni, hogy a fenti 4 egyenlőség közül akármelyik implikálja a többit, ezért a függetlenség definíciójául szolgálhat az egyetlen

$$\mathrm{P}(B|A) = \mathrm{P}(B)$$

egyenlőség is. Ha itt a baloldali feltételes valószínűséget hányadosként írjuk fel, akkor ezt kapjuk:

$$\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = \mathrm{P}(B)$$

amiből szorzással az A -ra, B -re nézve szimmetrikus

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\ \mathrm{P}(B)$$

egyenlőség adódik. Tehát a függetlenség szimmetrikus reláció: ha B független A -tól, akkor A is független B -től. Ezért függetlenség esetén azt mondhatjuk, hogy az események **függetlenek egymástól**. Könnyű belátni, hogy a

$$\mathrm{P}(A \cap B) \ = \ \mathrm{P}(A)\ \mathrm{P}(B)$$

egyenlőségből következnek a

$$\begin{aligned} \mathrm{P}(A \cap \overline{B}) &= \mathrm{P}(A)\mathrm{P}(\overline{B}) \\ \mathrm{P}(\overline{A} \cap B) &= \mathrm{P}(\overline{A})\mathrm{P}(B) \\ \mathrm{P}(\overline{A} \cap \overline{B}) &= \mathrm{P}(\overline{A})\mathrm{P}(\overline{B}) \end{aligned}$$

egyenlőségek is. Ezért a függetlenség definíciójául az egyetlen

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$

egyenlőség, vagy alábbi négy egyenlőség

$$\begin{aligned} \mathrm{P}(A \cap B) &= \mathrm{P}(A)\ \mathrm{P}(B) \\ \mathrm{P}(A \cap \overline{B}) &= \mathrm{P}(A)\ \mathrm{P}(\overline{B}) \\ \mathrm{P}(\overline{A} \cap B) &= \mathrm{P}(\overline{A})\ \mathrm{P}(B) \\ \mathrm{P}(\overline{A} \cap \overline{B}) &= \mathrm{P}(\overline{A})\ \mathrm{P}(\overline{B}) \end{aligned}$$

is szolgálhat.

Három esemény függetlensége: Három esemény függetlenségéhez már 8 egyenlőségnek kell teljesülni:

$$\begin{aligned}
\mathrm{P}(A_1 \cap A_2 \cap A_3) &= \mathrm{P}(A_1)\,\mathrm{P}(A_2)\,\mathrm{P}(A_3) \\
\mathrm{P}(A_1 \cap A_2 \cap \overline{A_3}) &= \mathrm{P}(A_1)\,\mathrm{P}(A_2)\,\mathrm{P}(\overline{A_3}) \\
\mathrm{P}(A_1 \cap \overline{A_2} \cap A_3) &= \mathrm{P}(A_1)\,\mathrm{P}(\overline{A_2})\,\mathrm{P}(A_3) \\
\mathrm{P}(A_1 \cap \overline{A_2} \cap \overline{A_3}) &= \mathrm{P}(A_1)\,\mathrm{P}(\overline{A_2})\,\mathrm{P}(\overline{A_3}) \\
\mathrm{P}(\overline{A_1} \cap A_2 \cap A_3) &= \mathrm{P}(\overline{A_1})\,\mathrm{P}(A_2)\,\mathrm{P}(A_3) \\
\mathrm{P}(\overline{A_1} \cap A_2 \cap \overline{A_3}) &= \mathrm{P}(\overline{A_1})\,\mathrm{P}(A_2)\,\mathrm{P}(\overline{A_3}) \\
\mathrm{P}(\overline{A_1} \cap \overline{A_2} \cap A_3) &= \mathrm{P}(\overline{A_1}1)\,\mathrm{P}(\overline{A_2})\,\mathrm{P}(A_3) \\
\mathrm{P}(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}) &= \mathrm{P}(\overline{A_1})\,\mathrm{P}(\overline{A_2})\,\mathrm{P}(\overline{A_3})
\end{aligned}$$

Több esemény függetlensége: Több, mondjuk n, esemény függetlensége esetén nem csak annak kell teljesülnie, hogy

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)\ldots\mathrm{P}(A_n),$$

hanem tetszőleges A_i -k helyett mindkét oldalon azok komplementerét véve is igaz kell hogy legyen az egyenlőség, például:

$$\mathrm{P}(\overline{A_1} \cap A_2 \cap \overline{A_3} \cap \overline{A_4} \cap \cdots \cap A_n) = \mathrm{P}(\overline{A_1})\mathrm{P}(A_2)\mathrm{P}(\overline{A_3})\mathrm{P}(\overline{A_4})\ldots\mathrm{P}(A_n)$$

Ilyen egyenletből 2^n darab van.

Feladat: Több vizsgálat jobb eredményt ad. Tegyük fel, hogy egy bizonyos ritka betegségben az embereknek csupán 1 ezreléke szenved. A betegséget egy olyan vizsgálattal lehet kimutatni, ami sajnos mindkét irányban tévedhet: beteg emberek esetén csak 0.8 a valószínűsége annak, hogy a vizsgálat jelzi a betegséget, egészséges emberek esetében pedig csak 0.9 a valószínűsége annak, hogy az egészséges embert egészségesnek jelzi. Barátomat nemrég többször is vizsgálták, és minden vizsgálat betegnek jelezte. A vizsgálatok számának függvényében adjuk meg, hogy mennyire jogos, hogy aggódik?

Megoldás: Tegyük fel, hogy a vizsgálatok száma n. Feltételezve, hogy a vizsgálatok függetlenek egymástól, igazak az alábbiak:

$$\mathrm{P}(\text{ betegnek diagnosztizálják } n \text{ -szer} \mid \text{beteg }) = (\ \mathrm{P}(\text{ betegnek diagnosztizálják egyszer} \mid \text{beteg })\ )^n$$

$$\mathrm{P}(\text{ betegnek diagnosztizálják } n \text{ -szer} \mid \text{nem beteg }) = (\ \mathrm{P}(\text{ betegnek diagnosztizálják egyszer} \mid \text{nem beteg })\ )^n$$

Ezeket felhasználva, – barátom esélyei így festenek:

$$\mathrm{P}(\text{ beteg} \mid \text{betegnek diagnosztizálják } n \text{ -szer })$$

$$= \frac{\mathrm{P}(\text{ beteg és betegnek diagnosztizálják } n \text{ -szer })}{\mathrm{P}(\text{ betegnek diagnosztizálják } n \text{ -szer })}$$

$$= \frac{\mathrm{P}(\text{ beteg })\cdot\mathrm{P}(\text{ betegnek diagnosztizálják } n \text{ -szer} \mid \text{beteg })}{\mathrm{P}(\text{ beteg })\cdot\ \mathrm{P}(\text{ betegnek diagnosztizálják } n \text{ -szer} \mid \text{beteg }) + \mathrm{P}(\text{ nem beteg })\cdot\ \mathrm{P}(\text{ betegnek diagnosztizálják } n \text{ -szer} \mid \text{nem beteg })}$$

$$= \frac{\mathrm{P}(\text{ beteg })\cdot(\ \mathrm{P}(\text{ betegnek diagnosztizálják egyszer} \mid \text{beteg })\ )^n}{\mathrm{P}(\text{ beteg })\cdot(\ \mathrm{P}(\text{ betegnek diagnosztizálják egyszer} \mid \text{beteg })\ )^n + \mathrm{P}(\text{ nem beteg })\cdot(\ \mathrm{P}(\text{ betegnek diagnosztizálják egyszer} \mid \text{nem beteg })\ )^n}$$

$$= \frac{0.001 \cdot 0.8^n}{0.001 \cdot 0.8^n\ +\ 0.999 \cdot 0.1^n}$$

Az alábbi táblázatban n függvényében adjuk meg a vizsgált feltételes valószínűség numerikus értékét:

n	P(beteg | betegnek diagnosztizálják n -szer)
1	0.008
2	0.060
3	0.339
4	0.804
5	0.970

Láthatjuk, hogy ha a teszt 5-szöri ismétlése mind pozitív eredményt ad, akkor már nagy a gond!

Megjegyzés: Felmerülhet valakiben a kérdés, hogy mi az esélye a betegségnek, ha n tesztből k jelez betegséget, de $n-k$ nem. A kérdésre a választ a binomiális eloszlás segítségével tudjuk majd megadni.

6.2. Valószínűségi változók függetlensége

Azt mondjuk, hogy az Y valószínűségi változó független az X valószínűségi változótól, ha minden x és y lehetséges érték esetén az $Y=y$ esemény független az $X=x$ eseménytől, azaz

$$\mathrm{P}(\ Y=y \mid X=x\ ) \quad = \quad \mathrm{P}(\ Y=y\ )$$

Ha itt a baloldali feltételes valószínűséget hányadosként írjuk fel, akkor ezt kapjuk:

$$\frac{\mathrm{P}(\ X=x\ \cap\ Y=y\ )}{\mathrm{P}(\ X=x\ )} \quad = \quad \mathrm{P}(\ Y=y\ )$$

amiből szorzással az A -ra, B -re nézve szimmetrikus

$$\mathrm{P}(\ X=x\ \cap\ Y=y\ ) \quad = \quad \mathrm{P}(\ X=x\ )\,\mathrm{P}(\ Y=y\ )$$

egyenlőség adódik. Tehát a függetlenség szimmetrikus reláció: ha Y független X -től, akkor X is független Y -tól. Ezért függetlenség esetén azt mondhatjuk, hogy a valószínűségi változók **függetlenek egymástól**.

6.3. Direktszorzat

1. Példa: Fiatal házaspár gyerekeinek száma és a nagyszülők száma függetlenek egymástól. Korábban vizsgáltuk az a problémát, melyben egy fiatal házaspárt véletlenszerűen választottunk, és tekintettük az alábbi két valószínűségi változót:

$$X = \text{ gyerekek száma}, \qquad Y = \text{ nagyszülők száma}$$

Az akkor mondott feltételek mellett meghatároztuk az $(X;Y)$ kétdimenziós valószínűségi változó eloszlását. $(X;Y)$ eloszlása mellett feltüntetjük X és Y eloszlását is:

y	P$(Y = y)$					
4	0.30	0.060	0.120	0.090	0.030	
3	0.40	0.080	0.160	0.120	0.040	
2	0.15	0.030	0.060	0.045	0.015	
1	0.10	0.020	0.040	0.030	0.010	
0	0.05	0.010	0.020	0.015	0.005	
		0.2	0.4	0.3	0.1	P$(X = x)$
		0	1	2	3	x

Vegyük észre, hogy $(X; Y)$ eloszlásának minden tagját szorzatként kaptuk meg X és Y eloszlásának megfelelő tagjaiból:

$$\mathrm{P}(X = x, Y = y) \;=\; \mathrm{P}(X = x)\,\mathrm{P}(Y = y)$$

Definíció: Ha egy kétdimenziós eloszlás minden tagja úgy képződik két egydimenziós eloszlás tagjaiból, hogy a megfelelő tagokat összeszorozzuk, akkor a síkbeli eloszlást a két egydimenziós eloszlás **direktszorzat**ának nevezzük.

6.4. Konvolúció

1. Példa: Gyerekek száma plusz nagyszülők száma. Korábban vizsgáltuk az a problémát, melyben egy fiatal házaspárt véletlenszerűen választottunk, és tekintettüük az alábbi két valószínűségi változót:

$$X = \text{ gyerekek száma}, \quad Y = \text{ nagyszülők száma}$$

Az akkor mondott feltételek mellett meghatároztuk az $(X; Y)$ kétdimenziós valószínűségi változó eloszlását:

y					
4	0.060	0.120	0.090	0.030	
3	0.080	0.160	0.120	0.040	
2	0.030	0.060	0.045	0.015	
1	0.020	0.040	0.030	0.010	
0	0.010	0.020	0.015	0.005	
	0	1	2	3	x

Az előző alpontban észrevettük, hogy $(X; Y)$ eloszlása X eloszlásából és Y eloszlásából direktszorzatként adódik:

y	$\mathrm{P}(Y=y)$					
4	0.30	0.060	0.120	0.090	0.030	
3	0.40	0.080	0.160	0.120	0.040	
2	0.15	0.030	0.060	0.045	0.015	
1	0.10	0.020	0.040	0.030	0.010	
0	0.05	0.010	0.020	0.015	0.005	
		0.2	0.4	0.3	0.1	$\mathrm{P}(X=x)$
		0	1	2	3	x

Ha – valamilyen rejtélyes okból a gyerekek számának és a nagyszülők számának az összege érdekel minket, akkor a $Z = X + Y$ valószínűségi változó eloszlával kell foglalkoznunk. Nyilván igazak az alábbiak:

$$
\begin{aligned}
&\mathrm{P}(Z=0) = 0.010\\
&\mathrm{P}(Z=1) = 0.040 = 0.020 + 0.020\\
&\mathrm{P}(Z=2) = 0.085 = 0.030 + 0.040 + 0.015\\
&\mathrm{P}(Z=3) = 0.175 = 0.080 + 0.060 + 0.030 + 0.005\\
&\mathrm{P}(Z=4) = 0.275 = 0.060 + 0.160 + 0.045 + 0.010\\
&\mathrm{P}(Z=5) = 0.255 = 0.120 + 0.120 + 0.015\\
&\mathrm{P}(Z=6) = 0.130 = 0.090 + 0.040\\
&\mathrm{P}(Z=7) = 0.030
\end{aligned}
$$

Ha Z lehetséges értékeit és a hozzájuk tapadó valószínűségeket táblázatba rendezzük, megkapjuk Z eloszlását:

z	0	1	2	3	4	5	6	7
$\mathrm{P}(Z=z)$	0.010	0.040	0.085	0.175	0.275	0.255	0.130	0.030

Z eloszlását egy egyszerű, **konvolúció**nak nevezett művelettel kaptuk meg X és Y eloszlásából: először vettük X eloszlásának és Y eloszlásának a direktszorzatát, majd a kapott síkbeli eloszlást a $z = x + y$ transzformációval a számegyenesre képeztük.

6.5. Gyakorló feladatok

1. Tegyük fel, hogy az év egy bizonyos napján Budapesten, New York-ban és Tokioban minden nap egymástól függetlenül esik az eső 0.6, 0.8, 0.5 valószínűségekkel, illetve nem esik 0.4, 0.2, 0.5 valószínűségekkel. Esőzés szempontjából a három városban 8 féle variáció lehetséges. Sorolja fel ezt a 8 variációt, és mind a 8 variációnak adja meg a valószínűségét!

2. *(Az előző feladat folytatása)*
   Hogyan egyszerűsödik az előző feladatban a számítás, ha az eső valószínűsége mindhárom városban ugyanannyi?

3. *(Példa páronként független, de összességében nem független eseményekre)*
   Egy 10 és egy 20 forintos érmével dobunk. Tekintsük az alábbi eseményeket:

   - $A =$ 10 forintos érme fejet ad
   - $B =$ 20 forintos érme fejet ad
   - $C =$ mindkét érmével írást dobok vagy mindkét érmével fejet dobok

   A és B nyilván függetlenek egymástól. Mutassuk meg, hogy

   (a) A és C függetlenek.
   (b) B és C függetlenek.
   (c) A és B és C nem függetlenek.

4. Egy piros és egy zöld kockával dobunk. Tekintsük az alábbi eseményeket:

   - $A =$ a dobott számok összege 7
   - $B =$ legalább az egyik kockán van hatos
   - $C =$ mindkét kockával páratlant dobok
   - $D =$ a két kockával különböző számokat dobok
   - $E =$ a zöld kockával 4-est dobok

   Válaszoljuk meg a következő kérdéseket:

   (a) Függetlenek-e egymástól az A és C események?
   (b) Kizáróak-e az A és C események?
   (c) Mennyi a B esemény valószínűsége?
   (d) Hogy viszonul egymáshoz A és D? Milyen következtetést vonhatunk le ebből a valószínűségeikre nézve? És a függetlenségekre nézve?
   (e) Függetlenek-e egymástól az A és E események?
   (f) Mindezek alapján mutassunk példát olyan eseményekre, amelyek
      i. függetlenek, de nem kizáróak,
      ii. kizáróak, de nem függetlenek.

7. Általános formulák kétdimenziós eloszlásokkal kapcsolatban

Amikor az X és Y valószínűségi változókból alkotunk egy kétdimenziós (X,Y) valószínűségi változót, kézenfekvő, hogy X,Y és (X,Y) súlyfüggvényei között kapcsolatokat fedezhetünk fel. A most következő általános formulákban x és y a súlyfüggvények változóit jelölik. Az (X,Y) kétdimenziós valószínűségi változó súlyfüggvényét $p(x,y)$, az X valószínűségi változó súlyfüggvényét $p_1(x)$, az Y valószínűségi változó súlyfüggvényét $p_2(y)$ jelöli. Emlékeztetünk a súlyfüggvények jelentésére:

$$p_1(x) = \mathrm{P}(\ X = x\ )$$

$$p_2(y) = \mathrm{P}(\ Y = y\ )$$

$$p(x,y) = \mathrm{P}(\ X = x,\ Y = y\ )$$

7.1. Összegzési szabály vetület eloszlásokkal kapcsolatban

Nyilvánvalóak az alábbi összegzési szabályok:

$$p_1(x) = \sum_y p(x,y)$$

$$p_2(y) = \sum_x p(x,y)$$

7.2. Szorzási szabály független koordináták esetére

Ha X és Y függetlenek, akkor

$$p(x,y) = p_1(x)\ p_2(y)$$

A $p(x,y)$ súlyfüggvényt (eloszlást) a $p_1(x)$ és $p_2(y)$ súlyfüggvények (eloszlások) **direktszorzat**ának nevezzük.

7.3. Osztási és szorzási szabályok feltételes súlyfüggvényekkel

Ha x lehetséges értéke X -nek, akkor az $X = x$ feltétel mellett Y feltételes súlyfüggvényét $p_{2|1}(y|x)$ -szel jelöljük. Tehát $p_{2|1}(y|x)$ annak a valószínűségét jelöli, hogy az $X = x$ feltétel mellett bekövetkezik az $Y = y$ esemény:

$$p_{2|1}(y|x) = \mathrm{P}(\ Y = y \mid X = x\ )$$

Nyilván teljesül az alábbi osztási szabály:

$$p_{2|1}(y|x) = \frac{p(x,y)}{p_1(x)}$$

Hasonlóképpen, ha y lehetséges értéke Y -nak, akkor az $Y = y$ feltétel mellett X feltételes súlyfüggvényét $p_{1|2}(x|y)$ -nal jelöljük, azaz

$$p_{1|2}(x|y) = \mathrm{P}(\ X = x \mid Y = y\ )$$

Nyilván teljesül az alábbi osztási szabály:

$$p_{1|2}(x|y) = \frac{p(x,y)}{p_2(y)}$$

A felírt osztási szabályok így írhatók át szorzási szabályokká:

$$p(x,y) = p_1(x)\ p_{2|1}(y|x)$$
$$p(x,y) = p_2(y)\ p_{1|2}(x|y)$$

7.4. Súlyfüggvények (diszkrét eloszlások) keverése

Az összegzési és a szorzási szabályok kombinálásával adódnak az alábbi formulák:

$$p_1(x) = \sum_y p_2(y)\ p_{1|2}(x|y)$$
$$p_2(y) = \sum_x p_1(x)\ p_{2|1}(y|x)$$

Ezek a formulák a korábban vett

$$\mathrm{P}(B) = \mathrm{P}(A_1)\cdot\mathrm{P}(B|A_1) + \mathrm{P}(A_2)\cdot\mathrm{P}(B|A_2) + \ldots \mathrm{P}(A_n)\cdot\mathrm{P}(B|A_n)$$

teljes valószínűség formulájának speciális esetei. Ezért hivatkozhatunk rájuk úgy is mint a **teljes valószínűség formulái kétdimenziós diszkrét valószínűségi változókra**. A jobboldali kifejezéseket egyfajta **keverés**nek is felfoghajuk: például a

$$p_2(y) = \sum_x p_1(x)\ p_{2|1}(y|x)$$

formula azt mutatja, hogy a $p_2(y)$ függvényértéket úgy kaphatjuk meg, hogy a a $p_{2|1}(y|x)$ függvényértékeket a $p_1(x)$ súlyfüggvény szerint keverjük. Tehát

- **az Y valószínűségi változó súlyfüggvénye (eloszlása) nem más, mint az Y feltételes súlyfüggvényeinek (feltételes eloszlásainak) keveréke az X súlyfüggvénye (eloszlása) szerint.**

Hasonlóképpen

- **az X valószínűségi változó súlyfüggvénye (eloszlása) nem más, mint az X feltételes súlyfüggvényeinek (feltételes eloszlásainak) keveréke az Y súlyfüggvénye (eloszlása) szerint.**

7.5. Súlyfüggvény (eloszlás) transzformációja egydimenzióban

Tegyük fel, hogy adott egy $y = t(x)$ függvény. Ha az X valószínűségi változó értékét behelyettesítjük az $y = t(x)$ függvénybe, akkor a kapott $t(X)$ érték egy új Y valószínűségi változót definiál:

$$Y = t(X)$$

Y súlyfüggvényének az értékét egy adott y helyen az X súlyfüggvényéből összegzéssel kapjuk meg:

$$r(y) = \sum_{x:\ t(x)=y} p(x)$$

A képlet azt fejez ki, hogy az $r(y)$ értéket úgy kaphatjuk meg, hogy megkeressük azokat az x helyeket, melyekre $t(x) = y$, és az ilyen x helyekhez tartozó $p(x)$ értékeket összeadjuk. (Tessék meggondolni, hogy ez miért van így!)

Azt azt mondjuk, hogy az új $r(y)$ súlyfüggvényt (eloszlást) az eredeti $p(x)$ súlyfüggvényből (eloszlásból) az $y = t(x)$ **transzformációval kaphatjuk meg**.

7.6. Súlyfüggvény (eloszlás) transzformációja kétdimenzióból egydimenzióba

Kétdimenziós esetben, ha adott $z = t(x, y)$ függvény, és az (X, Y) kétdimenziós valószínűségi változó értékét behelyettesítjük a $z = t(x, y)$ függvénybe, akkor a kapott $t(X, Y)$ érték egy új Z valószínűségi változót definiál: $Z = t(X, Y)$. A Z súlyfüggvényét (eloszlását) (X, Y) súlyfüggvényéből (eloszlásából)összegzéssel kapjuk meg:

$$r(z) = \sum_{(x,y):\ t(x,y)=z} p(x, y)$$

A képlet azt fejez ki, hogy az $r(z)$ értéket úgy kaphatjuk meg, hogy megkeressük azokat az (x, y) helyeket, melyekre $t(x, y) = z$, és az ilyen (x, y) helyekhez tartozó $p(x, y)$ értékeket összeadjuk. (Tessék meggondolni, hogy ez miért van így!)

Azt azt mondjuk, hogy az új súlyfüggvényt (eloszlást) az eredeti súlyfüggvényből (eloszlásból) a $z = t(x, y)$ **transzformációval kaphatjuk meg**.

7.7. Összeg súlyfüggvénye

Speciálisan, ha $z = x + y$, akkor $X + Y$ súlyfüggvényére a képlet így fest:

$$r(z) = \sum_{(x,y):\ x+y=z} p(x, y)$$

A képlet azt fejez ki, hogy az $r(z)$ értéket úgy kaphatjuk meg, hogy megkeressük azokat az (x, y) helyeket, melyekre $x + y = z$, és az ilyen (x, y) helyekhez tartozó $p(x, y)$ értékeket összeadjuk.

7.8. Konvolúció

Ha X és Y függetlenek, akkor $X + Y$ súlyfüggvénye az összegre vonatkozó előző képlet speciális esete:

$$r(z) = \sum_{(x,y):\ x+y=z} p_1(x) \cdot p_2(y) = \sum_{x} p_1(x) \cdot p_2(z - x) = \sum_{y} p_1(z - y) \cdot p_2(y)$$

A képlet azt fejez ki, hogy az $r(z)$ értéket úgy kaphatjuk meg, hogy megkeressük azokat az (x, y) helyeket, melyekre $x + y = z$, és az ilyen (x, y) helyekhez tartozó $p_1(x) \cdot p_2(y)$ szorzatértékeket összeadjuk.

Azt azt mondjuk, hogy az $r(z)$ súlyfüggvény (eloszlás) a $p_1(x)$ és $p_2(y)$ eloszlásokból **konvolúció**val adódik.

7.9. Gyakorló feladatok

Az alábbi számolós feladatok megoldásához előnyösen lehet használni az Excelt. Az Excel használata akkor lenne igazán hasznossá, ha táblázatok mérete lényegesen nagyobb lenne, mint a most következő feladatokban adott 5x4 -es táblázatok.

1. Töltse ki az alábbi táblázatot nemnegatív számokkal úgy, hogy azok összege 1 legyen, és tekintsen a táblázatra úgy, mintha a táblázat egy kétdimenziós $(X; Y)$ valószínűségi változó súlyfüggvényének a táblázata lenne.

y						
3						
2						
1						
0						
	0	1	2	3	4	x

(a) Adja meg X súlyfüggvényének értékeit az alábbi táblázatban:

0	1	2	3	4	x

(b) Adja meg Y súlyfüggvényének értékeit az alábbi táblázatban:

y	
3	
2	
1	
0	

A függőleges helyzetű "álló" táblázatot természetesen "le lehet fektetni", vízszintes helyzetbe:

0	1	2	3	y

(c) Adja meg a $Z = XY$ valószínűségi változó súlyfüggvényének értékeit egy megfelelő táblázatban!

(d) Adja meg a $Z = X + Y$ valószínűségi változó súlyfüggvényének értékeit egy megfelelő táblázatban!

2. *(Az 1. feladat folytatása.)* Töltse ki az alábbi táblázatot úgy, hogy egy az x értékhez tartozó oszlopba az $X = x$ feltétel melletti feltételes súlyfüggvény értékeit írja:

y						
3						
2						
1						
0						
	0	1	2	3	4	x

3. *(Az 1. feladat folytatása.)* Töltse ki az alábbi táblázatot úgy, hogy egy az y értékhez tartozó sorba az $Y = y$ feltétel melletti feltételes súlyfüggvény értékeit írja:

y						
3						
2						
1						
0						
	0	1	2	3	4	x

4. Töltse ki a

0	1	2	3	4	x

táblázatot pozitív számokkal úgy, hogy az összegük 1 legyen, és tekintsen a táblázatra úgy, mintha a táblázat egy X valószínűségi változó súlyfüggvényének a táblázata lenne.

Töltse ki a

y						
3						
2						
1						
0						
	0	1	2	3	4	x

táblázat minden oszlopát pozitív számokkal úgy, hogy minden oszlopban az összegük 1 legyen, és tekintsen a táblázatra úgy, mintha a táblázat minden oszlopa egy Y valószínűségi változó $X = x$ feltétel melletti feltételes súlyfüggvényének a táblázata lenne.

A felvett táblázatokra építve határozza meg az (X, Y) valószínűségi változó súlyfüggvényének a táblázatát:

y						
3						
2						
1						
0						
	0	1	2	3	4	x

5. Töltse ki az alábbi két táblázatot pozitív számokkal úgy, hogy mindkét táblázatban 1 legyen az összeg:

0	1	2	3	4	x

0	1	2	3	y

Határozza meg a két eloszlás

(a) direktszorzatát

(b) konvolúcióját!

6. Töltse ki az alábbi táblázatokat pozitív számokkal úgy, hogy minden táblázatban 1 legyen az összeg:

1. táblázat: az X valószínűségi változó súlyfüggvénye (eloszlása):

0	1	2	3	4	x

2. táblázat: az Y valószínűségi változó súlyfüggvénye (eloszlása az $X = 0$ feltétel mellett:

0	1	2	3	y

3. táblázat: az Y valószínűségi változó súlyfüggvénye (eloszlása) az $X = 1$ feltétel mellett:

0	1	2	3	y

4. táblázat: az Y valószínűségi változó súlyfüggvénye (eloszlása) az $X = 2$ feltétel mellett:

0	1	2	3	y

5. táblázat: az Y valószínűségi változó súlyfüggvénye (eloszlása) az $X = 3$ feltétel mellett:

0	1	2	3	y

6. táblázat: az Y valószínűségi változó súlyfüggvénye (eloszlása) az $X = 4$ feltétel mellett:

0	1	2	3	y

Határozza meg az Y valószínűségi változó súlyfüggvényét (eloszlását) úgy, hogy veszi a 2., 3., 4., 5., 6. táblázatban adott súlyfüggvények (eloszlások) keverékét az 1. táblázatban adott súlyfüggvény (eloszlás) szerint!

8. Vegyes feladatok eseményekre, valószínűségekre

1. Egy pakli francia kártyát (azaz 52 lapot, melyek között 4 ász van) véletlenszerűen négy játékosnak osztunk ki úgy, hogy mindenki kap 13 -13 lapot. Mi a valószínűsége, hogy mindenkinek jutott ász?

2. Egy diák három záróvizsgájára készül. Az első júniusban lesz, és ezen 90%-os eséllyel megy át. Ha ezen átment, akkor júliusban próbálhatja meg a második vizsgát, amely 80 % eséllyel lesz sikeres. Ha ezen is átment, akkor szeptemberben megy a harmadik vizsgára, ahol már csak 70% eséllyel megy át. Ellenben ha bármelyik vizsgája sikertelen, akkor csak egy év múlva lehet újra próbálkozni

   (a) Mi a valószínűsége annak, hogy az első évben átmegy mindhárom vizsgán?

   (b) Feltéve, hogy nem sikerült az első évben letenni a vizsgákat, mi a valószínűsége, hogy a második vizsgája volt sikertelen?

3. z A dobókockának 4 piros és 2 fehér oldala van, a B kockának pedig 2 piros és 4 fehér. Először feldobunk egy szabályos érmét. Ha fej, akkor a továbbiakban mindig az A kockával játszunk; ha írás, akkor pedig mindig a B kockával.

   (a) Mutassa meg, hogy a piros dobásának valószínűsége mindig $\frac{1}{2}$.

   (b) Ha az első két dobás piros, mi a valószínűsége, hogy a harmadik is piros?

   (c) Ha az első három dobás piros, akkor mi a valószínűsége, hogy az A kockát használjuk? (Csak a kocka felső lapját látjuk, a kocka többi oldalát nem.)

4. Egy dobozban 4 cédula van, három piros és egy kék. Kihúzunk egy cédulát, majd visszatesszük még három ugyanolyan színűvel együtt. Ezután ismét húzunk egy cédulát. Mi a valószínűsége annak, hogy

   (a) egyforma színű cédulákat húzunk?

   (b) pirosakat húzunk, feltéve, hogy egyforma színű cédulákat húzunk?

5. Információink szerint az A céggel kötött üzleteink 60%-a, a B céggel kötött üzletek 70%-a bizonyul kedvezőnek. Kettőjük közül a hamarabb jelentkező céggel rögtön két üzletet is kötünk. Feltehető, hogy $1/2$ valószínűséggel jelentkezik hamarabb A B -nél, és fordítva. Mi a valószínűsége, hogy

   (a) az első üzletkötés kedvező lesz?

   (b) mindkét üzletkötés javunkra válik?

   (c) lesz köztük rossz és jó üzlet is?

6. Ping-pongban az a nyertes, aki előbb éri el a 11 pontot, de legalább 2 pont különbség kell a nyeréshez (11 - 10 -nél folytatják két pont különbségig). Egy versenyen csak a nyertes kap pénzjutalmat: 1.000.000 Ft-ot. Két azonos képességű játékosnál a döntő szettben $10-9$ -es állásnál áramszünet lesz, nem lehet folytatni. Mi az igazságos osztozkodás a pénzen?

7. (a) Minden héten egy szelvénnyel játszunk az ötös lottón. Hány hétig kell ezt megtennünk, hogy legalább $\frac{1}{2}$ valószínűséggel legyen legalább kettes találatunk?

   (b) Mi a valószínűsége, hogy egy adott héten legalább kettes találatunk lesz, ha két függetlenül kitöltött szelvénnyel játszunk?

   (c) Mi a valószínűsége, hogy egy adott héten legalább kettes találatunk lesz, ha két olyan szelvénnyel játszunk, amelyeken nincsen egyforma szám? (Segítség: A 90 számot ossza háromfelé: 5+5 szám van a szelvényeken, 80 szám pedig nincs egyiken sem.)

8. Egy dobókockát az első hatosig, egy érmét az első fejig dobálunk. Mi a valószínűsége annak, hogy

   (a) a kockával pontosan annyit kell dobnunk, mint az érmével?

   (b) a kockával többet kell dobnunk, mint az érmével?

9. Kertemben 5 fehér és 7 barna tyúkot tartok. A fehér tyúkok átlagosan 3 naponként, a barnák 4 naponként tojnak egy-egy tojást.

   (a) Mi a valószínűsége annak, hogy tyúkjaim 24 óra alatt együttesen is csak egyetlen tojást produkálnak? Ma reggel a 12 tyúkból 2 beszökött az üres nyúlketrecbe, és rájuk záródott az ajtó. Mi a valószínűsége annak, hogy

   (b) mindkét tyúk barna volt?

   (c) az egyik fehér, a másik barna volt?

   (d) holnap reggel a nyúlketrecben tojás lesz?

   (e) a két tyúk egyike fehér, feltéve, hogy a ketrecben nem találok tojást?

   (f) a két tyúk egyike fehér, feltéve, hogy a másik barna, és a ketrecben nem találok tojást?

10. *(Kapkodó utazó.)* Egy utazó az íróasztalában, a nyolc fiók egyikében hagyta az útlevelét. Mielőtt a repülőtérre indulna, kapkodva próbálja megtalálni. A kapkodás miatt $0,1$ valószínűséggel akkor sem veszi észre az útlevelet, ha az éppen megnézett fiókban van.

    (a) Mi a valószínűsége, hogy nem találja meg az első 5 fiókban?

    (b) Ha nem találta meg az első 5 fiókban, mi a valószínűsége, hogy az útlevél nem is volt ezekben?

    (c) Ha nem találta meg az első 5 fiókban, mi a valószínűsége, hogy a hátralévő fiókokban megtalálja?

9. Nevezetes eloszlások

Az eloszlások táblázatai, súly- és eloszlásfüggvényeik grafikonjai a `Diszkret_eloszlasok.xls` fájlban láthatóak. Ezt a fájlt a tárgy honlapjáról nyithatják meg vagy tölthetik le.

9.1. Egyenletes eloszlások

9.1.1. Egyenletes eloszlás egydimenzióban

Mikor használjuk: Ha adott n darab szám, és ezek mindegyike ugyanakkora eséllyel bukkan fel.

Súlyfüggvény:

$$p(x) = \frac{1}{n} \quad \text{minden lehetséges } x \text{ -re}$$

Példa: Ilyen valószínűségi változóhoz jutunk, amikor szabályos dobókockával dobunk, és

$$X = \text{ a kockán felül lévő szám}$$

$$p(x) = \frac{1}{6} \quad (x = 1, 2, 3, 4, 5, 6)$$

9.1.2. Egyenletes eloszlás kétdimenzióban

Mikor használjuk: Ha adott n darab pont a síkon (emlékeztetőül: egy pont a síkon egy számpárt jelent), és ezek a pontok mindegyike ugyanakkora eséllyel bukkan fel.

Súlyfüggvény:

$$p(x, y) = \frac{1}{n} \quad \text{minden lehetséges } (x, y) \text{ pontra}$$

Példa: Ilyen valószínűségi változóhoz jutunk, amikor egy piros és egy zöld dobókockával dobunk, és

$$X = \text{ a piros kockán felül lévő szám}$$

$$Y = \text{ a zöld kockán felül lévő szám}$$

$$p(x, y) = \frac{1}{36} \quad (x = 1, 2, 3, 4, 5, 6; \quad y = 1, 2, 3, 4, 5, 6)$$

9.1.3. Egyenletes eloszlás r -dimenzióban

Mikor használjuk: Ha adott n darab pont az r -dimenziós térben (emlékeztetőül: egy r -dimenziós pont egy r elemű sorozatot jelent), és mindegyik pont ugyanakkora eséllyel bukkan fel.

Súlyfüggvény:

$$p(x_1, x_2, \dots x_r) = \frac{1}{n} \qquad \text{minden lehetséges } (x_1, x_2, \dots x_r) \text{ pontra}$$

Példa: Ilyen 3 -dimenziós valószínűségi változóhoz jutunk, amikor egy piros, egy zöld és egy kék dobókockával dobunk, és

$$X = \text{ a piros kockán felül lévő szám}$$

$$Y = \text{ a zöld kockán felül lévő szám}$$

$$Z = \text{ a kék kockán felül lévő szám}$$

$$p(x, y, z) = \frac{1}{216} \quad (x = 1, 2, 3, 4, 5, 6; \quad y = 1, 2, 3, 4, 5, 6; \quad z = 1, 2, 3, 4, 5, 6)$$

9.2. Hipergeometrikus eloszlások

9.2.1. Hipergeometrikus eloszlás

Súlyfüggvény:

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad \text{ha} \quad x \geq 0, \quad x \geq n - N + M, \quad x \leq n, \quad x \leq M$$

Mikor használjuk: Ha N darab golyó van egy ládában, közülük M darab piros, a többi $N - M$ darab fehér, és visszatevés **nélkül** húzunk n -szer, akkor az

$$X = \text{ahányszor pirosat húzunk}$$

valószínűségi változó ilyen eloszlást követ.

Paraméterek jelentése:

$$N = \text{ az összes golyók száma a dobozban a dobozban a húzások megkezdése előtt}$$

$$M = \text{ a piros golyók száma a dobozban a húzások megkezdése előtt}$$

$$n = \text{ a húzások száma}$$

A súlyfüggvény képletének levezetése: Az N golyó közül n darabot kihúzni $\binom{N}{n}$ féle képpen lehet, ennyi az összes kombinációk száma. Az $X = x$ esemény azt jelenti, hogy a kihúzott n golyó között x darab piros és $n - x$ darab fehér van. Ezért azt kell meghatároznunk, hogy hány olyan kombináció van, amiben x darab piros és $n - x$ darab fehér golyó van. Az x darab piros golyó a M darab piros közül $\binom{M}{x}$ féleképpen kerülhet ki. Az $n - x$ fehér golyó az $N - M$ darab fehér közül $\binom{N-M}{n-x}$ féleképpen kerülhet ki. Bármely piros kombinációt bármelyik fehér kombinációval össze lehet párosítani. Ezért az „ x darab piros és $n-x$ darab fehér" eseményre nézve kedvező kombinációk száma a két szám szorzata:

$$\binom{M}{x}\binom{N-M}{n-x}$$

Ezért a keresett súlyfüggvény érték:

$$p(x) = \mathrm{P}(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Példa: Ha az ötös lottón egy szelvénnyel játszom, és X jelöli a találataim számát, akkor a 2, 3, 4, illetve 5 találat valószínűsége:

$$\mathrm{P}(X=2) = \frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}} \qquad \mathrm{P}(X=3) = \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} \qquad \mathrm{P}(X=4) = \frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}} \qquad \mathrm{P}(X=5) = \frac{\binom{5}{5}\binom{85}{0}}{\binom{90}{5}}$$

azaz

$$p(x) = \frac{\binom{5}{x}\binom{85}{5-x}}{\binom{90}{5}} \qquad \text{ha} \quad 0 \le \; x \le 90$$

Excel-függvények:

$$p(x) = \texttt{HYPGEOM.DIST(x;n;M;N;FALSE)} = \texttt{HIPERGEOM.ELOSZLÁS(x;n;M;N;HAMIS)}$$

$$F(x) = \texttt{HYPGEOM.DIST(x;n;M;N;TRUE)} = \texttt{HIPERGEOM.ELOSZLÁS(x;n;M;N;IGAZ)}$$

1. Példa: Hány szarvas él az erdőben? Egy erdőben ismeretlen számú szarvas él. A számuk becslése céljából 60 szarvast piros festékkel megjelölünk, majd néhány hét eltelte után megszámoljuk, hogy 40 véletlenszerűen választott szarvas között hány megjelöltre bukkanunk. Tegyük fel, hogy 15 -re. Mit mondhatunk ezek alapján a szarvasok ismeretlen N számáról?

Megoldás: Átfogalmazzuk a feladatot erdőben szabadon élő szarvasok helyett dobozba zárt golyókra. Igaz, így kevésbé izgalmas a probléma, de könnyebben elképzelhető.

2. Példa: Hány golyó van a dobozban? Egy dobozban ismeretlen számú fehér golyó van, melyek közül 60 -at pirosra festünk, majd jól összekeverjük a golyókat, és 40 -szer húzunk visszatevés nélkül. Tegyük fel, hogy a kihúzott golyók között 15 piros akad. Mit mondhatunk ezek alapján a golyók ismeretlen N számáról?

1. Megoldás: Az összes golyók között a pirosak aránya $60 : N$. A kihúzott golyók között a pirosak aránya $15 : 40$. Feltételezve, hogy ez a két arány körülbelül egyenlő, a

$$60 : N \;\approx\; 15 : 40$$

közelítő egyenletet kapjuk, amiből az jön ki, hogy

$$N \;\approx\; 60 * \frac{40}{15} \;=\; 160$$

2. Megoldás: Ha X -szel jelöljük azt a valószínűségi változót, hogy hányszor húzunk piros golyót, akkor X hipergeometrikus eloszlást követ N, 60, 40 paraméterekkel. Az ismeretlen N paraméter néhány értéke mellett – a hipergeometrikus eloszlás képlete alapján – kiszámoltuk a P(15 piros) valószínűséget, és az értékeket táblázatba foglaltuk:

N	P(15 piros)
100	0.00
120	0.02
140	0.11
160	0.15
180	0.12
200	0.08
220	0.04
240	0.02
260	0.01
280	0.01
300	0.00

Vegyük észre, hogy $N = 160$ esetén a valószínűség értéke 0.15, de $N \leq 100$ vagy $N \geq 300$ esetén – két tizedesre kerekítve – 0.00, ami igazából azt jelenti, hogy a a valószínűség értéke 0.005 -nél is kisebb. Ezért úgy gondolhatjuk, hogy a golyók (szarvasok) száma 160 körül van. A táblázatban szereplő valószínűségekből arra következtetünk, hogy a golyók (szarvasok) száma 100 és 300 között van.

Az, hogy a golyók (szarvasok) száma 100 és 300 között van, a mi tippünk. Egyáltalán nem biztos, hogy ez így is van. Lehet, hogy a golyók (szarvasok) száma 100 -nál kevesebb vagy 300 -nál több, és ezért jól kinevetnek minket a szarvasok, és azt mondják, hogy "Bibibí, rosszul tippeltetek, mi többen vagy kevesebben vagyunk annál, mint ahogy azt ti tippeltétek". De megnyugtató érzést ad nekünk, hogy annak az esélye, hogy kinevetnek, kicsúfolnak minket a szarvasok, garantáltan kisebb, mint 0.005.

Egy másik, kevésbé szégyenlős ember nagyobb rizikót is felvállalhat. Abból a tényből, hogy 15 piros golyó van a 40 kihúzott között, szűkebb határt is mondhat a golyók (szarvasok) számára. Például azt, hogy 120 és 240 között van a golyók (szarvasok) száma. Az, hogy a golyók (szarvasok) száma 120 és 240 között van, ez az ő tippje. Egyáltalán nem biztos, hogy igaza van. Lehet, hogy a golyók (szarvasok) száma 120 -nál kevesebb vagy 240 -nél több, és ezért jól kinevetik őt a szarvasok, és azt mondják, hogy "Bibibí, rosszul tippeltél, mi többen vagy kevesebben vagyunk annál, mint ahogy azt te tippelted". Ezt az embert megnyugtatja az a tény, hogy annak az esélye, hogy kinevetik és kicsúfolják a szarvasok, garantáltan kisebb, mint 0.025. Az ő esetében érvényes 0.025 nagyobb, mint a mi esetünkben érvényes 0.005. Ez az ára annak, hogy ő szűkebb határokkal tippel, mint mi.

Tehát azt látjuk, hogy ilyen vagy olyan alsó és felső határokkal tippelve a golyók (szarvasok) számára, annak az esélye, hogy kinevetnek és kicsúfolnak minket a szarvasok, legfeljebb ennyi vagy annyi:

- a 100 – 300 -as tipp esetén legfeljebb 0.005
- a 120 – 240 -es tipp esetén legfeljebb 0.025

Azt a nyilvánvaló tényt, hogy szűkebb határok esetén a szarvasoknak nagyobb esélyük lehet a nevetésre, csúfolásra, mint tágabb határok esetén, a valószínűségszámítás elmélete alapján számszerű összefüggéssel tudtuk finomítani. Ezt a számszerű összefüggést a fentebb megadott táblázatból olvastuk ki (a táblázat által megengededett pontosság erejéig). A valószínűségszámítás elmélete – többek között - ilyesmire tanít meg minket.

9.2.2. Polihipergeometrikus eloszlás

Súlyfüggvény:

$$p(x;y) = \frac{\binom{M_1}{x}\binom{M_2}{y}\binom{N-M_1-M_2}{n-x-y}}{\binom{N}{n}}$$

ha

$$0 \le x \le \min(n, M_1),$$
$$0 \le y \le \min(n, M_2),$$
$$0 \le n - x - y \le \min(n, N - M_1 - M_2) \le n$$

Mikor használjuk: Ha egy dobozban N golyó van, melyek közül M_1 darab piros, M_2 darab zöld, $N - M_1 - M_2$ darab fehér, és n -szer húzunk a dobozból **visszatevés nélkül**, akkor az a kétdimenziós valószínűségi változó, melynek koordinátái

$$X = \text{ahányszor piros golyót húzunk}$$

$$Y = \text{ahányszor zöld golyót húzunk}$$

ilyen eloszlást követ.

A súlyfüggvény képletének levezetése: Az N golyó közül n darabot kihúzni $\binom{N}{n}$ féle képpen lehet, ennyi az összes kombinációk száma. Az $X = x, Y = y$ esemény azt jelenti, hogy a kihúzott n golyó között x darab piros, x darab zöld és $n - x - y$ darab fehér van. Ezért azt kell meghatároznunk, hogy hány olyan kombináció van, amiben x darab piros, y darab zöld és $n-x-y$ darab fehér golyó van. Az x darab piros golyó az M_1 darab piros közül $\binom{M_1}{x}$ féleképpen kerülhet ki. Az y darab zöld golyó az M_2 darab zöld közül $\binom{M_2}{y}$ féleképpen kerülhet ki. Az $n - x - y$ fehér golyó az $N - M_1 - M_2$ darab fehér közül $\binom{N-M_1-M_2}{n-x-y}$ féleképpen kerülhet ki. Bármely piros kombinációt bármelyik zöld és bármelyik fehér kombinációval össze lehet rakni. Ezért az x darab , y darab zöld és $n - x$ darab fehér" eseményre nézve kedvező kombinációk száma e három szám szorzata:

$$\binom{M_1}{x}\binom{M_2}{y}\binom{N-M_1-M_2}{n-x-y}$$

Ezért a keresett súlyfüggvény érték:

$$p(x, y) = \mathrm{P}\left(X = x, y = y\right) = \frac{\binom{M_1}{x}\binom{M_2}{y}\binom{N-M_1-M_2}{n-x-y}}{\binom{N}{n}}$$

Megjegyzés: Nyilvánvaló, hogy ha a dobozból való húzásokat visszatevéssel végezzük, akkor az (X, Y) valószínűségi változó polinomiális eloszlást követ $p_1 = \frac{M_1}{N}$, $p_2 = \frac{M_2}{N}$ paraméterekkel. A polinomiális eloszlásról néhány oldallal később lesz szó.

9.2.3. Polihipergeometrikus eloszlás, r -dimenziós *(Extra tananyag)*

Súlyfüggvény:

$$p(x_1, x_2, \ldots, x_r) = \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\ldots\binom{M_r}{x_r}\binom{N-M_1-M_2-\ldots-M_r}{n-x_1-x_2-\ldots-x_r}}{\binom{N}{n}}$$

ha

$$0 \le x_i \le \min(n, M_i) \qquad (1 \le i \le r)$$
$$0 \le n - x_1 - x_2 - \ldots - x_r \le \min(n, N - M_1 - M_2 - \ldots - M_r)$$

Mikor használjuk: Ha egy dobozban N golyó van, melyek közül M_1 darab piros, M_2 darab zöld, és így tovább M_r darab lila golyó van, továbbá $N - M_1 - M_2 - \dots M_r$ darab fehér, és n -szer húzunk a dobozból **visszatevés nélkül**, akkor az az n -dimenziós valószínűségi változó, melynek koordinátái

$$X_1 = \text{ahányszor piros golyót húzunk}$$

$$X_2 = \text{ahányszor zöld golyót húzunk}$$

$$\vdots$$

$$X_r = \text{ahányszor lila golyót húzunk}$$

ilyen eloszlást követ.

Megjegyzés: Nyilvánvaló, hogy ha a dobozból való húzásokat visszatevéssel végezzük, akkor az $(X_1, X_2, \dots, X_r)$ valószínűségi változó polinomiális eloszlást követ

$$p_1 = \frac{M_1}{N} \qquad p_2 = \frac{M_2}{N} \qquad \dots \qquad p_r = \frac{M_r}{N}$$

paraméterekkel.

9.3. Binomiális eloszlás és társai

9.3.1. Binomiális eloszlás

Súlyfüggvény:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{ha} \quad x = 0, 1, 2, \dots, n$$

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Egy p valószínűségű eseményre n kísérletet végzünk. Az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahányszor bekövetkezik az esemény az } n \text{ kísérlet során}$$

   Paraméterek jelentése:

   $$n = \text{ a kísérletek száma}$$

   $$p = \text{ az esemény valószínűsége}$$

2. **Több eseménnyel kapcsolatban:** n darab független, külön-külön p valószínűségű esemény kapcsán az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahány esemény bekövetkezik az } n \text{ esemény közül}$$

   Paraméterek jelentése:

   $$n = \text{ az események száma}$$

   $$p = \text{ az események közös valószínűség értéke}$$

3. **Golyók húzása dobozból:** Ha N darab golyó van egy ládában, közülük M darab piros, a többi $N - M$ darab fehér, és n -szer húzunk **visszatevéssel**, akkor az valószínűségi változó ilyen eloszlást követ:

$$X = \text{ahányszor pirosat húzunk}$$

Paraméterek jelentése:

$$n = \text{ a húzások száma}$$
$$p = M/N = \text{ piros húzásának valószínűsége minden egyes húzásnál}$$

A súlyfüggvény képletének levezetése: A levezetést egy példán keresztül mutajuk be. E célból tegyük fel, hogy egy 10 tagú társaságból mindenki, a többiektől függetlenül, p valószínűséggel megy el vasárnap kirándulni. Jelölje X a kirándulók számát. Most azt kérdezzük, hogy mennyi az X súlyfügvényének az értéke az $x = 3$ helyen, azaz mennyi annak az események a valószínűsége, hogy $X = 3$, vagyis hogy pontosan 3 ember megy el kirándulni:

$$p(3) = \mathrm{P}\,(\,X = 3\,) = \,?$$

Válasz: Az, hogy 3 ember megy el kirándulni, bekövetkezhet például úgy, hogy a társaság 3 legöregebb tagja elmegy, a többi 7 pedig nem. Ennek a valószínűsége, a független eseményekre vonatkozó szorzási szabály miatt nyilván:

$$p\,p\,p\ \ (1-p)\,(1-p)\,(1-p)\,(1-p)\,(1-p)\,(1-p)\,(1-p)\ =$$
$$= \ p^3\,(1-p)^7$$

Az, hogy 3 ember megy el kirándulni, természetesen másképpen is lehetséges: a társaság akármelyik 3 tagja elmehet, a többi 7 pedig nem. Akármelyik 3 tag esetében annak a valószínűsége, hogy ők elmennek, a többi 7 pedig nem, ugyancsak

$$p^3\,(1-p)^7$$

Ezért, ha nem fontos számunkra, hogy ki az a 3 ember, aki elmegy, csak az fontos számunkra, hogy pontosan 3 ember menjen el (márpedig az $X = 3$ esemény esetében ez a helyzet), akkor ennek az eseménynek a valószínűségét úgy kapjuk meg, hogy a $p^3\,(1-p)^7$ közös valószínűség értéket megszorozzuk a 3 elemű kombinációk számával, vagyis $\binom{10}{3}$ -mal:

$$p(3) = \mathrm{P}\,(\,X = 3\,) = \binom{10}{3}\,p^3\,(1-p)^7$$

Világos, hogy ha a társaság nem 10 tagú, hanem n tagú, akkor annak a valószínűsége, hogy az n ember közül pontosan x ember megy el kirándulni:

$$p(x) = \mathrm{P}\,(\,X = x\,) = \binom{n}{x} p^x (1-p)^{n-x}$$

Példa: Ha dobókockával 20-szor dobok, és a dobott hatosok számát X jelöli, akkor a súlyfüggvény:

$$p(x) = \binom{20}{x} p^x (1-p)^{20-x} \qquad (x = 0, 1, \ldots, n)$$

A pontosan 3 hatos valószínűsége:

$$\mathrm{P}(X = 3) = \binom{20}{3} (1/6)^3 (5/6)^{17}$$

Megjegyzés: A binomiális eloszlást a **visszatevéses**, a hipergeometrikus eloszlást a **visszatevés nélküli** húzások esetén kell használni.

Excel-függvények:

$$p(x) = \texttt{BINOM.DIST(x;n;p;FALSE)} = \texttt{BINOM.ELOSZLÁS(x;n;p;HAMIS)}$$
$$F(x) = \texttt{BINOM.DIST(x;n;p;TRUE)} = \texttt{BINOM.ELOSZLÁS(x;n;p;IGAZ)}$$

Megjegyzés: Egyes Excel verziókban az eloszlások nevében a . (pont) elhagyható, – például –

$$\texttt{BINOM.DIST} \text{ helyett } \texttt{BINOMDIST}$$

írható.

Megjegyzés: Hipergeometrikus eloszlás $\approx$ binomiális eloszlás. Kis ügyeskedéssel be lehet látni, hogy akármilyen x pozitív egész szám esetén az

$$M \to \infty\,, \quad N \to \infty\,, \quad M/N \to p$$

feltételek mellett

$$\lim \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

Ez azt jelenti, hogy ha x -hez képest M is és N is nagy, és az M/N arány közelítőleg p, akkor a hipergeometrikus eloszlás x -ik tagja közelítőleg egyenlő a binomiális eloszlás x -ik tagjával:

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x}$$

Az egzakt levezetést az "ügyeskedésekkel" az Olvasóra bízzuk. Cserébe egy heurisztikus magyarázatot adunk színes golyók segítségével:

Heurisztikus magyarázat: Tegyünk N golyót egy dobozba úgy, hogy közülül M piros, a többi fehér. Húzzunk a dobozból n -szer visszatevés nélkül. Annak a esélye, hogy pontosan x -szer húzunk pirosat – mint tudjuk – a hipergeometrikus eloszlás képlete szerint

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Most tegyünk N golyót egy másik dobozba ugyanúgy, mint az előbb, de most visszatevéssel húzzunk a dobozból n -szer. Annak a esélye, hogy pontosan x -szer húzunk pirosat – mint tudjuk – a binomiális eloszlás képlete szerint

$$\binom{n}{x} \left(\frac{M}{N}\right)^x \left(1 - \frac{M}{N}\right)^{n-x}$$

Ha x -hez képest M is és N is nagy, akkor a kihúzott piros golyók számára nézve nincs sok hatása annak, hogy visszatevés nélkül vagy visszatevéssel húzúnk, ezért az x darab piros golyó esélye így is úgy is kb ugyanannyi:

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} \left(\frac{M}{N}\right)^x \left(1 - \frac{M}{N}\right)^{n-x}$$

Mivel feltettük, hogy az M/N arány közelítőleg p, a jobboldali kifejezést kicserélhetjük így:

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x}$$

Ezek után talán kedve támad az Olvasónak, hogy az egzakt levezetést "kiügyeskedje". Hajrá!

1. Feladat: Vajon, mindenki le tud ülni? 400 hallgató mindegyike egymástól függetlenül 0.6 valószínűséggel jár órára. A teremben 250 db szék van. Mi a valószínűsége annak, hogy aki elmegy órára, mind le tud ülni egy-egy székre?

Megoldás: Először vegyük észre, hogy az órát látogató diákok X száma valószínűségi változó, mely – a diákok habitusa miatt – binomiális eloszlást követ $n = 400$ és $p = 0.6$ paraméterekkel. Az az esemény, hogy mindenkinek, aki elmegy órára, jut szék, azt jelenti, hogy $X \le 250$. Ennek az eseménynek a $\mathrm{P}(X \le 250)$ valószínűségét az eloszlásfüggvénynek a 250 helyen felvett értéke adja meg, ami Excellel így számolható:

$$F(250) = \texttt{BINOM.DIST(250 ; 400 ; 0.6 ; TRUE)} =$$

$$= \texttt{BINOM.ELOSZLÁS(250 ; 400 ; 0.6 ; IGAZ)} = 0.86$$

Tehát 0.86 valószínűséggel csak 250 vagy kevesebb hallgató meg órára, ezért 250 széken remekül elférnek. 0.14 valószínűséggel 250 -nél több hallgató megy órára, ezért ilyenkor lesz, aki nem tud székre ülni.

2. Feladat: Biztos, hogy mindenki le tud ülni? Hány szék kell ahhoz, hogy biztosan (1 valószínűséggel, 100% biztonsággal) jusson szék mindenkinek, aki elmegy órára?

Megoldás: Ahhoz, hogy biztosan jusson szék mindenkinek, aki elmegy órára, nyilván 400 vagy több székre van szükség.

3. Feladat: Spóroljunk a székekkel! Hány szék kell ahhoz, hogy legalább 0.99 legyen a valószínűsége annak (99% legyen a biztonsága annak), hogy aki elmegy órára, mind le tud ülni egy-egy székre?

Megoldás: Ahhoz, hogy legalább 0.99 legyen a valószínűsége annak, hogy mindenkinek, aki elmegy órára, jusson szék, az kell, hogy a székek x száma eleget tegyen az alábbi egyenlőtlenségnek:

$$F(x) = \mathrm{P}(X \le x) \ge 0.99$$

Ezért a

$$\texttt{BINOM.DIST(x ; 400 ; 0.6 ; TRUE)} =$$

$$= \texttt{BINOM.ELOSZLÁS(x ; 400 ; 0.6 ; IGAZ)}$$

függvényre készítettünk egy táblázatot:

x	$F(x)$
260	0.9824
261	0.9864
262	0.9897
263	0.9922
264	0.9942
265	0.9957
266	0.9968
267	0.9977
268	0.9984
269	0.9988
270	0.9992

amiből kiolvashatuk, hogy x értéke, vagyis a székek száma legalább 263, akkor $F(x) = \mathrm{P}(X \le x) \ge 0.99$.

1. Megjegyzés: Vegyük észre, hogy ha a 100% biztonság helyett megelégszünk 99% biztonsággal, akkor a székeknek több mint 1/3 -át megspórolhatjuk! Ha valaki a 99% biztonsággal még nem elégedett, akkor tegyen be a 263 székhez még 7-et, és a biztonság – a táblázat utolsó sora szeribnt – megnő 99.9%-re.

2. Megjegyzés: Ilyen jellegű feladatok adódnak, amikor valamilyen rendszer, például egy telefonközpont, kapacitását kell megtervezni úgy, hogy a rendszer nagy valószínűséggel el tudja látni a véletlenszerűen felmerülő igényeket.

4. Feladat: A gubanc valószínűsége. Tegyük fel, hogy egy repülőgépen 200 ülés van az utasok számára, de – bízván abban, hogy néhány utas betegség, forgalmi dugó, stb. miatt lemarad a járatról – az extra profit reményében 203 jegyet adnak el. Feltéve, hogy minden utas a többitől függetlenül 0.05 valószínűséggel marad le a járatról, számoljuk ki annak a valószínűségét, hogy a beszálláskor "gubanc támad", mert 200 -nál több utas jelenik meg?

Megoldás: A járathoz időben odaérkező utasok száma nyilván binomiális eloszlást követ $n = 203$ és $p = 0.95$ paraméterekkel. Ezért annak a valószínűsége, hogy 200 -nál több utas jelenik meg a beszálláskor, vagyis előáll a "gubanc", egyenlő

$$1 - \texttt{BINOM.DIST(200 ; 203 ; 0.95 ; TRUE)} =$$

$$= 1 - \texttt{BINOM.ELOSZLÁS(200 ; 203 ; 9.5 ; IGAZ)} = 0.0021$$

5. Feladat: Hány extra repülőjegyet adjunk el? Az előző feladatban kapott valószínűség értéke elég kicsi ahhoz, hogy a repülőtársaság felvállalja a gubanc következményeit. Nézzük meg, hogyan függ a gubanc valószínűsége az eladott extra jegyek számától! Gondoljuk meg, hogy a p paraméter függvényében hány extra jegyet adjon el a társaság, ha nem szeretne túl gyakran gubanc-ot!

Megoldás: Excelben egyszerűen megszerkeszthető egy olyan táblázat, ami a gubanc valószínűségét mutatja p és az extra jegyek számának függvényében:

P(gubanc)	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	p
1	0.1326	0.0172	0.0022	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
2	0.3992	0.0865	0.0154	0.0025	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	
3	0.6685	0.2265	0.0555	0.0113	0.0021	0.0003	0.0001	0.0000	0.0000	0.0000	
4	0.8507	0.4159	0.1368	0.0354	0.0078	0.0015	0.0003	0.0000	0.0000	0.0000	
5	0.9437	0.6091	0.2613	0.0844	0.0224	0.0051	0.0010	0.0002	0.0000	0.0000	
6	0.9818	0.7675	0.4144	0.1647	0.0522	0.0140	0.0033	0.0007	0.0001	0.0000	
7	0.9948	0.8763	0.5721	0.2751	0.1036	0.0323	0.0086	0.0021	0.0004	0.0001	
8	0.9987	0.9407	0.7120	0.4056	0.1794	0.0647	0.0198	0.0053	0.0013	0.0003	
9	0.9997	0.9741	0.8211	0.5414	0.2781	0.1152	0.0400	0.0120	0.0032	0.0008	
10	0.9999	0.9897	0.8971	0.6675	0.3926	0.1856	0.0728	0.0244	0.0072	0.0019	
extra jegyek száma											

A táblázatból kiolvashatjuk, hogy ha a gubanc valószínűségét 0.005 alatt akarjuk tartani, akkor

$p = 0.01$	esetén	0
$p = 0.02$	esetén	0
$p = 0.03$	esetén	1
$p = 0.04$	esetén	2
$p = 0.05$	esetén	3
$p = 0.06$	esetén	4
$p = 0.07$	esetén	6
$p = 0.08$	esetén	7
$p = 0.09$	esetén	9
$p = 0.10$	esetén	10

darab extra jegy adható el.

Feladat: Amikor a tesztek különböző eredményeket adnak! Tegyük fel, hogy egy bizonyos ritka betegségben az embereknek csupán 1 ezreléke szenved. A betegséget egy olyan vizsgálattal lehet kimutatni, ami sajnos mindkét irányban tévedhet: beteg emberek esetén csak 0.8 a valószínűsége annak, hogy a vizsgálat jelzi a betegséget, egészséges emberek esetében pedig csak 0.9 a valószínűsége annak, hogy az egészséges embert egészségesnek jelzi. Barátomat nemrég n -szer is vizsgálták, és k vizsgálat betegnek jelezte, $n-k$ pedig nem. Mi a valószínűsége annak, hogy barátom beteg?

Megoldás: Feltételezve, hogy a vizsgálatok függetlenek egymástól, igazak az alábbiak:

$$\text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)} = \binom{n}{k} 0.8^k (1-0.8)^{n-k} = \text{BINOM.DIST(k ; n ; 0.8 ; FALSE)}$$

$$\text{P(betegnek diagnosztizálják } k \text{ -szor | nem beteg)} = \binom{n}{k} 0.1^k (1-0.1)^{n-k} = \text{BINOM.DIST(k ; n ; 0.1 ; FALSE)}$$

Ezeket felhasználva – barátom esélyei így festenek:

$$\text{P(beteg } | \text{ betegnek diagnosztizálják } k \text{ -szor)} =$$

$$= \frac{\text{P(beteg és betegnek diagnosztizálják } k \text{ -szor)}}{\text{P(betegnek diagnosztizálják } k \text{ -szor)}}$$

$$= \frac{\text{P(beteg)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)}}{\text{P(beteg)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)} + \text{P(nem beteg)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | nem beteg)}}$$

$$= \frac{0.001 \cdot \binom{n}{x} 0.8^x (1-0.8)^{n-x}}{0.001 \cdot \binom{n}{x} 0.8^x (1-0.8)^{n-x} + 0.999 \cdot \binom{n}{x} 0.1^x (1-0.1)^{n-x}} =$$

$$= \frac{0.001 \cdot \text{BINOM.DIST(k ; n ; 0.8 ; FALSE)}}{0.001 \cdot \text{BINOM.DIST(k ; n ; 0.8 ; FALSE)} + 0.999 \cdot \text{BINOM.DIST(k ; n ; 0.1 ; FALSE)}}$$

Az alábbi táblázat a P(beteg | betegnek diagnosztizálják k -szor) feltételes valószínűség numerikus értékeit tartalmazza a $0 < k \leq n \leq 5$ esetekre:

	5	4	3	2	1	k
1					0.008	
2				0.060	0.002	
3			0.339	0.014	0.000	
4		0.804	0.102	0.003	0.000	
5	0.970	0.477	0.025	0.001	0.000	
n						

Ez a másik a táblázat pedig az $n = 10$ és $k = 9, 8, 7, 6, 5, 4$ esetekről szól:

	9	8	7	6	5	4	k
$n = 10$	1.000	0.999	0.958	0.390	0.017	0.000	

Érdemes elgondolkodni a numerikus értékek jelentésén!

Megjegyzés: A binomiális elolszlás $n = 1$ mellett vett speciális esete lesz az indikátor eloszlás, többdimenziós általánosítása pedig a polinomiális eloszlás.

9.3.2. Indikátor eloszlás

Súlyfüggvény:

$$p(x) = \begin{cases} p & \text{ha} \quad x = 1 \\ 1-p & \text{ha} \quad x = 0 \end{cases}$$

Mikor használjuk: Egy eseménnyel kapcsolatban bizonyos gondolatmenetkben hasznos, ha az esemény bekövetkezését egy kétértékű valószínűségi változóval kódoljuk: az 1 az esemény bekövetkezését, a 0 az esemény be nem következését jelenti:

$$X = \begin{cases} 1 & \text{ha az esemény bekövetkezik} \\ 0 & \text{ha az esemény nem következik be} \end{cases}$$

Ezt a valószínűségi változót az esemény **indikátorá**nak nevezzük. Ha például több eseménnyel kapcsolatban azt nézzük, hogy közülük hány következik be, akkor ez a valószínűségi változó az egyes események indikátorainak az összege. A binomiális eloszlással kapcsolatos számításoknál hasznos, hogy indikátorok összegeként binomiális eloszlású valószínűségi változót tudunk előállítani.

Paraméter jelentése:

$$p = \text{az esemény valószínűsége}$$

9.3.3. Binomiális eloszlás számsorozaton

Ha az $a_0, a_1, a_2, \ldots, a_n$ sorozat egy számtani sorozatot alkot, akkor

$$p(a_k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, 2, \ldots, n)$$

képlettel definiált eloszlás neve: **binomiális eloszlás a** $\{a_0, a_1, a_2, \ldots, a_n\}$ **halmazon**.

Nyilvánvaló, hogy ha egy eseménnyel kapcsolatban n kísérletet végzünk, akkor az esemény relatív gyakorisága binomiális eloszlást követ a

$$\{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{k}{n} \ldots, \frac{n}{n}\}$$

halmazon.

9.3.4. Polinomiális eloszlás

Súlyfüggvény:

$$p(x; y) = \frac{n!}{x!\, y!\, (n-x-y)!} \ (p_1)^x \ (p_2)^y \ (1 - p_1 - p_2)^{n-x-y}$$

ha

$$0 \le x \le n$$

$$0 \le y \le n$$

$$0 \le n - x - y \le n$$

Mikor használjuk: Ha A és B egymást kizáró események, és n kísérletet végzünk velük kapcsolatban, akkor az a kétdimenziós valószínűségi változó, melynek koordinátái

$$X = \text{ahányszor bekövekezik az } A \text{ esemény}$$

$$Y = \text{ahányszor bekövekezik a } B \text{ esemény}$$

ilyen eloszlást követ.

Paraméterek jelentése:

$$p_1 = \text{ az } A \text{ esemény valószínűsége}$$

$$p_2 = \text{ a } B \text{ esemény valószínűsége}$$

A súlyfüggvény képletének levezetése: A levezetést egy példán keresztül mutajuk be. E célból tegyük fel, hogy egy 10 tagú társaságból mindenki, a többiektől függetlenül, vasárnap p_1 valószínűséggel megy el kirándulni, p_2 valószínűséggel pedig moziba megy. Jelölje X a kirándulók számát, Y pedig a moziba menők számát. Most azt kérdezzük, hogy mennyi az (X, Y) súlyfügvényének az értéke az $x = 3, y = 2$ helyen, azaz mennyi annak az események a valószínűsége, hogy $X = 3$, és $Y = 2$, vagyis hogy pontosan 3 ember megy el kirándulni, és pontosan 2 megy moziba:

$$p(3, 2) = \mathrm{P}\,(\,X = 3\,,\,Y = 2\,) = ?$$

Válasz: Az, hogy 3 ember megy el kirándulni, és 2 megy moziba, bekövetkezhet például úgy, hogy a társaság 3 legöregebb tagja kirándulni megy, a 2 legfiatalabb moziba megy, a többi 5 pedig sem nem kirándul, sem nem megy moziba. Ennek a valószínűsége, a független eseményekre vonatkozó szorzási szabály miatt nyilván:

$$p_1\,p_1\,p_1\;\;p_2\,p_2\;\;(1-p_1-p_2)\,(1-p_1-p_2)\,(1-p_1-p_2)\,(1-p_1-p_2)\,(1-p_1-p_2)\;=$$

$$=\;p_1^3\,p_2^2\,(1-p_1-p_2)^5$$

Az, hogy 3 ember megy el kirándulni, és 2 pedig moziba természetesen másképpen is lehetséges: a társaság akármelyik 3 tagja mehet kirándulni, akármelyik másik 2 tagja mehet moziba, és a többi 5 pedig se ide se oda nem megy. Akármelyik 3 kirándulós és 2 mozizós ember esetében annak a valószínűsége, hogy ők mennek kirándulni, illetve moziba, és a többi 5 pedig se ide se oda, ugyancsak

$$p_1^3\,p_2^2\,(1-p_1-p_2)^5$$

Ezért, ha nem fontos számunkra, hogy ki az a 3 ember, aki kirándulni megy, és ki az a 2 ember, aki moziba megy, csak az fontos számunkra, hogy pontosan 3 ember menjen kirándulni és pontosan 2 pedig moziba (márpedig az

$$X = 3\;,\;Y = 2$$

esemény esetében ez a helyzet), akkor ennek az eseménynek a valószínűségét úgy kapjuk meg, hogy a

$$p_1^3\,p_2^2\,(1-p_1-p_2)^5$$

közös valószínűség értéket megszorozzuk a 3 kiránduló és a 2 mozizó kiválasztási lehetőségeinek számával, vagyis

$$\binom{10}{3}\,\binom{7}{2}$$

-vel:

$$p(3, 2) = \mathrm{P}\,(\,X = 3\,,\,Y = 2\,) = \binom{10}{3}\,\binom{7}{2}\,p_1^3\,p_2^2\,(1-p_1-p_2)^5$$

A binomiális együtthatókat faktoriálisokkal kifejtve, egyszerűsítés után azh eredményt így is írhatjuk:

$$p(3, 2) = \mathrm{P}\,(\,X = 3\,,\,Y = 2\,) = \frac{10!}{3!\;2!\;5!}\;p_1^3\,p_2^2\,(1-p_1-p_2)^5$$

Világos, hogy ha a társaság nem 10 tagú, hanem n tagú, akkor annak a valószínűsége, hogy az n ember közül pontosan x ember megy el kirándulni, és y ember mozizni:

$$p(x,y) = \mathrm{P}\,(\,X = x\,,\,Y = y\,) = \frac{n!}{x!\;y!\;(n-x-y)!}\;p_1^x\,p_2^y\,(1-x-y)^{n-x-y}$$

Példa: Ha szabályos dobókockával 20 -szor dobunk, és

$X =$ ahányszor 1 -est dobunk

$Y =$ ahányszor 2 -est vagy 3 -ast dobunk

akkor az (X,Y) valószínűségi változó kétdimenziós polinomiális eloszlást követ $n = 20$, $p_1 = \frac{1}{6}$, $p_2 = \frac{2}{6}$ paraméterekkel. A súlyfüggvény képlete:

$$p(x,y) = \frac{20!}{x!\;y!\;(20-x-y)!}\;\left(\frac{1}{6}\right)^x\;\left(\frac{2}{6}\right)^y\;\left(1-\frac{1}{6}-\frac{2}{6}\right)^{(20-x-y)}$$

Például az $X = 3$ és $Y = 5$ esemény valószínűsége:

$$\mathrm{P}(\;X = 3\;,\;Y = 5\;) = \frac{20!}{3!\;5!\;(20-3-5)!}\;\left(\frac{1}{6}\right)^3\;\left(\frac{2}{6}\right)^5\;\left(1-\frac{1}{6}-\frac{2}{6}\right)^{(20-3-5)}$$

Megjegyzés: Vegyük észre, hogy az itt bevezetett polinomiális eloszlás a binomiális eloszlás általánosítása egydimenzióról kétdimenzióra.

9.3.5. Polinomiális eloszlás, r -dimenziós *(Extra tananyag)*

Súlyfüggvény:

$$p(x_1,x_2,\ldots,x_r) = \frac{n!}{x_1!\;x_2!\ldots x_r!\;(n-x_1-x_2-\ldots-x_r)!}\;(p_1)^{x_1}\;(p_2)^{x_2}\;\ldots(p_r)^{x_r}\;(1-p_1-p_2-\ldots-p_r)^{(n-x_1-x_2-\ldots-x_r)}$$

ha

$$0 \le x_i \le n \qquad (1 \le i \le r)$$

$$0 \le n - x_1 - x_2 - \ldots - x_r \le n$$

Mikor használjuk: Ha $A_1, A_2, \ldots, A_r$ egymást kizáró események, és n kísérletet végzünk velük kapcsolatban, akkor az az r -dimenziós valószínűségi változó, melynek koordinátái

$X_1 =$ ahányszor bekövekezik az A_1 esemény

$X_2 =$ ahányszor bekövekezik az A_2 esemény

$\vdots$

$X_r =$ ahányszor bekövekezik az A_r esemény

Paraméterek jelentése:

$p_1 =$ az A_1 esemény valószínűsége

$$p_2 = \text{ az } A_2 \text{ esemény valószínűsége}$$
$$\vdots$$
$$p_r = \text{ az } A_r \text{ esemény valószínűsége}$$

Példa: Ha szabályos dobókockával 20 -szor dobunk, és

$$X_1 = \text{ ahányszor 1 -est dobunk}$$
$$X_2 = \text{ ahányszor 2 -est vagy 3 -ast dobunk}$$
$$X_3 = \text{ ahányszor 4 -est dobunk}$$

akkor az (X_1, X_2, X_3) valószínűségi változó 3-dimenziós polinomiális eloszlást követ

$$n = 20 \qquad p_1 = \frac{1}{6} \qquad p_2 = \frac{2}{6} \qquad p_3 = \frac{1}{6}$$

paraméterekkel.

Megjegyzés: Vegyük észre, hogy az itt bevezetett polinomiáliss eloszlás a binomiális eloszlás általánosítása egydimenzióról r -dimenzióra.

9.4. Geometriai eloszlások és társaik

9.4.1. Geometriai eloszlás (optimista)

Súlyfüggvény:

$$p(x) = (1-p)^{x-1}p \qquad \text{ha} \quad x = 1, 2, \ldots$$

Megjegyzés: Vegyük észre, hogy a súlyfüggvény egy mértani sorozat, melynek első tagja p és kvóciense $q = 1 - p$.

Eloszlásfüggvény:

$$F(x) = 1 - (1-p)^x \quad (x = 1, 2, \ldots)$$

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:**Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahányadik kísérletnél először következik be az esemény}$$

   Paraméter jelentése:

   $$p = \text{ az esemény valószínűsége}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén

$$X = \text{ahányadik esemény az első olyan, ami bekövetkezik}$$

Paraméter jelentése:

$$p = \text{ az események közös valószínűség értéke}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X jelölöli annak a kísérletnek a sorszámát, amikor az A esemény először következik be, akkor az $X = x$ esemény azt jelenti, hogy az első $x - 1$ kísérlet során az A nem következett be, de az x -ik kísérletnél A bekövetkezett. Ezért – a független eseményekre vonatkozó szorzási szabály felhasználva – ezt kapjuk:

$$p(x) = \mathrm{P}\left(X = x\right) = (1 - p)^{x-1}p$$

1. Feladat: Vadászat az első nyúlig. Vadász barátom hétköznapokon (hétfőtől péntekig) a rosszabbik puskáját használja. Ezzel a puskával minden lövése a többitől függetlenül csak 0.05 valószínűséggel teríti le a célba vett nyulat. Hétvégén (szombaton és vasárnap) a jobb puskájával lövöldöz. Ezzel a puskával minden lövése a többitől függetlenül 0.4 valószínűséggel teríti le a célba vett nyulat. Minden nap addig lövöldöz a nyulakra, míg sikerül egyet leterítenie.

Első kérdés: Azt mesélik, hogy egy bizonyos napon, melyről nem tudjuk, hogy a hét melyik napja volt, 7 lövést adott le a nyúlvadászaton, hogy teljesítse aznapi feladatát. Ez a bizonyos nap vajon hétköznap volt, vagy hétvége?

Második kérdés: Hogyan lehet következtetni a napi lövések számából arra, hogy a lövések napja vajon hétköznap volt, vagy hétvége?

Megoldás: Tekintsünk egy hétköznapot. A nyúl leterítéséhez szükséges lövések száma optimista geometriai eloszlást követ 0.1 paraméterrel. Ezért egy hétköznapon annak a valószínűsége, hogy k lövés dördül,

$$\mathrm{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) = 0.05\,(1 - 0.05)^{k-1}$$

Egy hétvégi napokon annak a valószínűsége, hogy k lövés dördül,

$$\mathrm{P}(\ k \text{ lövés} \mid \text{hétvégi nap}\ ) = 0.4\,(1 - 0.4)^{k-1}$$

Mivel a napok 5/7-e hétköznap, 2/7-e hétvégi nap, annak a valószínűsége, hogy egy véletlenszerűen választott napon k lövés dördül,

$$\frac{5}{7} \cdot 0.05\,(1 - 0.05)^{k-1} + \frac{2}{7} \cdot 0.4\,(1 - 0.4)^{k-1}$$

Ha egy véletlenszerűen választott napon k lövés dördül, akkor annak a valószínűsége, hogy ez a nap hétköznap volt:

$$\mathrm{P}(\ \text{hétköznap} \mid k \text{ lövés}\ ) = \frac{\frac{5}{7} \cdot 0.05\,(1 - 0.05)^{k-1}}{\frac{5}{7} \cdot 0.05\,(1 - 0.05)^{k-1} + \frac{2}{7} \cdot 0.4\,(1 - 0.4)^{k-1}}$$

Annak a valószínűsége pedig, hogy hétvégi nap volt:

$$\mathrm{P}(\ \text{hétvégi nap} \mid k \text{ lövés}\ ) = \frac{\frac{2}{7} \cdot 0.4\,(1 - 0.4)^{k-1}}{\frac{5}{7} \cdot 0.1\,(1 - 0.1)^{k-1} + \frac{2}{7} \cdot 0.4\,(1 - 0.4)^{k-1}}$$

Excellel könnyű kiszámolni a képletek numerikus értékét. Íme a táblázat:

k	P(hétköznap | k lövés)	P(hétvégi nap | k lövés)
1	0.24	0.76
2	0.33	0.67
3	0.44	0.56
4	0.55	0.45
5	0.66	0.34
6	0.76	0.24
7	0.83	0.17
8	0.89	0.11
9	0.93	0.07
10	0.95	0.05

Válasz az első kérdésre: A táblázatból kiolvassuk, hogy $k = 7$ esetén a hétköznap valószínűsége 0.83 , a hétvégi nap valószínűsége 0.17

Válasz a második kérdésre: A táblázat azt is mutatja, hogy $k = 1, 2, 3$ esetén a hétvégi nap a valószínűbb, más lövészszám esetén a hétköznap a valószínűbb.

9.4.2. Geometriai eloszlás (pesszimista)

Súlyfüggvény:

$$p(x) = (1-p)^x p \qquad \text{ha} \quad x = 0, 1, 2, \ldots$$

Megjegyzés: Vegyük észre, hogy a súlyfüggvény egy mértani sorozat, melynek első tagja p és kvóciense $q = 1-p$. Az optimista geometriai eloszláshoz képest ez az eloszlás annyiban más, hogy most a lehetséges értékek az $x = 0, 1, \ldots$ számok, míg az optimista esetben a lehetséges értékek az $x = 1, 2, \ldots$ számok. Szemléletesen úgy fogalmazhatunk, hogy a pesszimista geometriai eloszlás súlyfüggvényének az ábrája úgy származtatható az optimistából, hogy a súlyfügvény grafikonját egy egységgel balra toljuk.

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahányszor nem következi be az esemény az első bekövetkezés előtt}$$

   Más terminológiát használva:

   $$X = \text{ahány kudarc van az első siker előtt}$$

   Paraméter jelentése:

   $$p = \text{ a siker valószínűsége}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahány kudarcos esemény van az első sikeres előtt}$$

   Paraméter jelentése:

   $$p = \text{ az események közös valószínűség értéke}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X azt jelölöli, hogy az A esemény első bekövetkezése előtt hányszor nem következett be, akkor az $X = x$ esemény azt jelenti, hogy az első x kísérlet során az A nem következett be, de az $(x+1)$ -ik kísérletnél A bekövetkezett. Ezért – a független eseményekre vonatkozó szorzási szabály felhasználva – ezt kapjuk:

$$p(x) = \mathrm{P}(X = x) = (1-p)^x p$$

Megjegyzés: Az "optimista", "pesszimista" jelzők használatát az indokolja, hogy az optimista esetben **sikeres** kísérletre vadászunk, a pesszimista esetben pedig **a kudarcok** számát számoljuk.

Megjegyzés: A geometriai elolszlások általánosításai a negatív binomiális eloszlások.

9.4.3. Negatív binomiális eloszlás (optimista)

Súlyfüggvény:

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \text{ha } x = r, r+1, r+2, \ldots$$

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahányadik kísérletnél } r \text{ -edszer következik be az esemény}$$

   Más terminlógiát használva:

   $$X = \text{ahányadik kísérletnél adódik az } r \text{ -ik sikeres esemény}$$

   Paraméterek jelentése:

   $$p = \text{ az esemény valószínűsége}$$
   $$r = \text{ ahányadik bekövetkezésre (sikerre) vadászunk}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahányadik esemény az } r \text{ -ik olyan, ami bekövetkezik}$$

   Más terminlógiát használva:

   $$X = \text{ahányadik az } r \text{ -ik sikeres esemény}$$

   Paraméterek jelentése:

   $$p = \text{ az események közös valószínűség értéke}$$
   $$r = \text{ ahányadik bekövetkezésre (sikerre) vadászunk}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X jelölöli annak a kísérletnek a sorszámát, amikor az A esemény r -edszer következik be, akkor az $X = x$ esemény azt jelenti, hogy az első $(x-1)$ kísérlet során az A esemény $(r-1)$ -szer következett be, és az A esemény az x -ik kísérletnél is bekövetkezik. Annak a valószínűsége, hogy az első $(x-1)$ kísérlet során az A esemény $(r-1)$ -szer következik be, a binomiális eloszlás képőlete szerint

$$\binom{x-1}{r-1} p^{r-1}(1-p)^{(x-1)-(r-1)} = \binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}$$

Mivel az x -ik kísérletnél az A esemény bekövetkezik, a keresett valószínűséget úgy kapjuk meg, hogy ezt a valószínűséget még megszorozzuk p -vel:

$$p(x) = \mathrm{P}(X = x) = \binom{x-1}{r-1} p^{r-1}(1-p)^{x-r)}\, p = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}$$

Excel-függvények:

$p(x) =$ `NEGBINOM.DIST(x-r;r;p;FALSE)` $=$ `NEGBIN.ELOSZLÁS(x-r;r;p;HAMIS)`

$p(x) =$ `NEGBINOM.DIST(x-r;r;p;TRUE)` $=$ `NEGBIN.ELOSZLÁS(x-r;r;p;IGAZ)`

1. Feladat: Nyúlvadászat rafináltabb módon. Vadász barátom hétköznapokon (hétfőtől péntekig) a rosszabbik puskáját használja. Ezzel a puskával minden lövése a többitől függetlenül 0.3 valószínűséggel teríti le a célba vett nyulat. Hétköznapokon addig lövöldöz a nyulakra, míg síkerül négyet leterítenie. Hétvégén (szombaton és vasárnap) az ünnepi puskájával lövöldöz. Az ünnepi puskával minden lövése a többitől függetlenül 0.8 valószínűséggel teríti le a célba vett nyulat. Ünnepnapokon addig lövöldöz a nyulakra, míg síkerül hatot leterítenie.

Első kérdés: Azt mesélik, hogy egy bizonyos napon, melyről nem tudjuk, hogy a hét melyik napja volt, 7 lövést adott le a nyúlvadászaton, hogy teljesítse aznapi feladatát. Ez a bizonyos nap vajon hétköznap volt, vagy hétvége?

Második kérdés: Hogyan lehet következtetni a napi lövések számából arra, hogy a lövések napja vajon hétköznap volt, vagy hétvége?

Megoldás: Tekintsünk egy hétköznapot, amikor négy nyúl leterítése a feladat. Az ehhez szükséges lövések száma optimista negatív binomiális eloszlást követ 4 és 0.3 paraméterekkel. Ezért egy hétköznapon annak a valószínűsége, hogy k lövés dördül,

P(k lövés | hétköznap) $=$ `NEGBINOM.DIST( k-4 ; 4 ; 0.3 ; FALSE )`

Egy hétvégi napon, amikor hat nyúl leterítése a feladat, a lövések száma optimista negatív binomiális eloszlást követ 6 és 0.8 paraméterekkel. Ezért hétvégi napokon annak a valószínűsége, hogy k lövés dördül,

P(k lövés | hétvégi napon) $=$ `NEGBINOM.DIST( k-6 ; 6 ; 0.8 ; FALSE )`

Mivel a napok 5/7-e hétköznap, 2/7-e hétvégi nap, annak a valószínűsége, hogy egy véletlenszerűen választott napon k lövés dördül,

$$\frac{5}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) + \frac{2}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ )$$

Ha egy véletlenszerűen választott napon k lövés dördül, akkor annak a valószínűsége, hogy ez a nap hétköznap volt:

$$\mathrm{P}(\ \text{hétköznap} \mid k \text{ lövés}\ ) = \frac{\frac{5}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétköznap}\ )}{\frac{5}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) + \frac{2}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ )}$$

Annak a valószínűsége pedig, hogy hétvégi nap volt:

$$\mathrm{P}(\text{ hétvégi nap} \mid k \text{ lövés }) = \frac{\frac{2}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétvégi napon })}{\frac{5}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétköznap }) + \frac{2}{7} \cdot \mathrm{P}(\ k \text{ lövés} \mid \text{hétvégi napon })}$$

Tekintve, hogy a képletek számlálóiban, illetve nevezőiben szereplő feltételes valószínűségek numerikus értékét – a fentebb adott képletekkel – Excelben ki lehet számolni a szóbajövő k értékekre, a keresett feltételes valószínűségekre is tudunk táblázatot készíteni:

k	P(hétköznap | k lövés)	P(hétvégi nap | k lövés)
4	1.00	0.00
5	1.00	0.00
6	0.27	0.73
7	0.31	0.69
8	0.44	0.56
9	0.62	0.38
10	0.79	0.21
11	0.90	0.10
12	0.96	0.04
13	0.99	0.01
14	0.99	0.01
15	1.00	0.00

Válasz az első kérdésre: A táblázatból kiolvassuk, hogy $k = 7$ esetén a hétköznap valószínűsége 0.31 , a hétvégi nap valószínűsége 0.69

Válasz a második kérdésre: A táblázat azt is mutatja, hogy $k = 6, 7, 8$ esetén a hétvégi nap a valószínűbb, más lövészszám esetén a hétköznap a valószínűbb.

9.4.4. Negatív binomiális eloszlás (pesszimista)

Súlyfüggvény:

$$p(x) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad \text{ha} \quad x = 0, 1, 2, \ldots \quad \text{(kombinatórikus formula)}$$

$$p(x) = \binom{-r}{x} p^r \left(-(1 - p) \right)^x \quad \text{ha} \quad x = 0, 1, 2, \ldots \quad \text{(analitikus formula)}$$

Megjegyzés: Vegyük észre, hogy az optimista negatív binomiális eloszláshoz képest ez az eloszlás annyiban más, hogy most a lehetséges értékek az $x = 0, 1, \ldots$ számok, míg az optimista esetben a lehetséges értékek az $x = r, r + 1, r + 2, \ldots$ számok. Szemléletesen úgy fogalmazhatunk, hogy a pesszimista negatív binomiális eloszlás grafikonja ábrája az optimistából úgy származtatható, hogy a grafikont r egységgel balra toljuk.

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

$$X = \text{ahány kudarc van az } r \text{ -ik siker előtt}$$

   Paraméterek jelentése:

$$p = \text{ az esemény valószínűsége}$$

$$r = \text{ ahányadik bekövetkezésre vadászunk}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az alábbi valószínűségi változó követ ilyen eloszlást:

$$X = \text{ahány kudarcos esemény van az } r \text{ -ik sikeres előtt}$$

   Paraméterek jelentése:

$$p = \text{ az események közös valószínűség értéke}$$

$$r = \text{ ahányadik bekövetkezésre vadászunk}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X jelölöli azt, hogy az A esemény az r -edik bekövetkezése előtt hányszor nem következett be, akkor az $X = x$ esemény azt jelenti, hogy az első $(x + r - 1)$ kísérlet során az A esemény $(r - 1)$ -szer következett be, és az A esemény az $(x + r)$ -ik kísérletnél is bekövetkezik. Annak a valószínűsége, hogy az első $(x + r - 1)$ kísérlet során az A esemény $(r - 1)$ -szer következik be, a binomiális eloszlás képőlete szerint

$$\binom{x+r-1}{r-1} p^{r-1} (1-p)^{(x+r-1)-(r-1)} = \binom{x+r-1}{r-1} p^{r-1} (1-p)^{x}$$

Mivel az $(x + r)$ -ik kísérletnél az A esemény bekövetkezik, a keresett valószínűséget úgy kapjuk meg, hogy ezt a valószínűséget még megszorozzuk p -vel:

$$p(x) = \mathrm{P}(X = x) = \binom{x+r-1}{r-1} p^{r-1} (1-p)^{x}\, p = \binom{x+r-1}{r-1} p^{r} (1-p)^{x}$$

Excel-függvények:

$$p(x) = \texttt{NEGBINOM.DIST(x;r;p;FALSE)} = \texttt{NEGBIN.ELOSZLÁS(x;r;p;HAMIS)}$$

$$p(x) = \texttt{NEGBINOMM.DIST(x;r;p;TRUE)} = \texttt{NEGBIN.ELOSZLÁS(x;r;p;IGAZ)}$$

Megjegyzés: Vegyük észre, hogy az (optimista, pesszimista) negatív binomiális eloszlások az (optimista, pesszimista) geometriai eloszlások általánosításai. $r = 1$ esetén az (optimista, pesszimista) negatív binomiális eloszlás a megfelelő geometriai eloszlásba egyszerűsödik.

9.5. Poisson eloszlás

9.5.1. Poisson eloszlás egydimenzióban

Súlyfüggvény:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad \text{ha} \quad x = 0, 1, 2, \ldots$$

Mikor használjuk: Sok független, külön-külön kis valószínűségű esemény kapcsán az alábbi valószínűségi változó ilyen eloszlást követ:

$$X = \text{ahány esemény bekövetkezik}$$

Paraméter jelentése:

$$\lambda = \text{ ahány esemény } \textit{általában, átlagosan} \text{ bekövetkezik}$$

azaz

$$\lambda = \text{ az eloszlás várható értéke (ezt a fogalmat a következő fejezetben tanuljuk)}$$

azaz

$$\lambda = \text{ az események valószínűségeinek összege}$$

Megjegyzés: Ha az események az események száma n, és az események egyforma valószínűségűek, és ez a közös érték p, akkor

$$\lambda = np$$

Excel-függvények:

$$p(x) = \texttt{POISSON(x;}\lambda\texttt{;FALSE)} = \texttt{POISSON(x;}\lambda\texttt{;HAMIS)}$$
$$F(x) = \texttt{POISSON(x;n;p;TRUE)} = \texttt{POISSON(x;n;p;IGAZ)}$$

Megjegyzés: Be lehet látni, hogy ha λ egy fix pozitív szám, akkor akármilyen x pozitív egész szám esetén az

$$n \to \infty\,, \qquad p \to 0\,, \qquad n \cdot p \to \lambda$$

feltételek mellett

$$\lim \binom{n}{x} p^x (1-p)^{n-x} \;=\; \frac{\lambda^x}{x!} e^{-\lambda}$$

Ez azt jelenti, hogy ha n nagy és p pici úgy, hogy szorzatuk körülbelül λ, akkor a binomiális eloszlás x -ik tagja közelítőleg egyenlő a Poisson eloszlás x -ik tagjával:

$$\binom{n}{x} p^x (1-p)^{n-x} \;\approx\; \frac{\lambda^x}{x!} e^{-\lambda}$$

Igazából ennek a ténynek a birtokában mondhatjuk azt, amit fentebb mondjunk: sok független, külön-külön kis valószínűségű esemény kapcsán az

$$X = \text{ahány esemény bekövetkezik}$$

valószínűségi változó eloszlását binomiális eloszlás helyett vehetjük Poisson eloszlásnak.

1. feladat: Hány hal lesz az *ÖREG* halász hálójában? Történt egyszer, hogy egy nagy tó partján álldogálva nézegettem, ahogy a sok apró halacska egymással és az öreg, alig-alig látó halásszal mit sem törődve össze-vissza úszkált. A halász éppen a hálóját készült kiemelni, amikor kedvesen így szólt hozzám: "Fiatalember! Ha eltalálod, hogy hány hal lesz a hálóban, meghívlak vacsorára". Szeretem a sült halat, és éhes is voltam. Csak annyit kérdeztem az öreg halásztól, hogy milyen gyakran szokott üres lenni a háló. Ő erre azt felelte:

"Sok éves tapasztalatom szerint mondhatom neked, hogy az eseteknek a 6 % -ában üres a háló!"

Kicsit gondolkodtam, számolgattam, és 2 halra tippeltem. Hogyan gondolkodtam, miért éppen 2 halra tippeltem?

Megoldás: Nyilván azt kellett kigondolnom, hogy a hálóban hány hal a legvalószínűbb. Így okoskodtam: A hálóban lévő halak száma – mint valószínűségi változó – Poisson eloszlást követ, hiszen

- a tóban *sok* a hal, és
- a sok hal mindegyike egymással mit sem törődve össze-vissza úszkál, ezért *egymástól függetlenül* kerülnek vagy nem kerülnek a hálóba, továbbá
- minden hal esetén az az esemény, hogy ő a hálóba kerül (ez az esemmény a hal számára felettébb szomorú), *kis valószínűségű*, hiszen ez a valószínűség a háló (akármilyen) méretének és a tó méretének a hányadosa, ami nyilván pici.

Mivel – sok éves tapasztalata alapján – a halász elárulta, hogy az eseteknek a 6 % -ában üres a háló, a 0 darab hal valószínűségét 0.06 -nak vettem, vagyis gondolatban felállatottam a

$$\frac{\lambda^0}{0!}\mathrm{e}^{-\lambda} = 0.06 \qquad \text{azaz} \qquad \mathrm{e}^{-\lambda} = 0.06$$

egyenletet, aminek megoldása $\lambda = -\ln(0.06) = 2.81$.

A 2.81 paraméterű Poisson eloszlás – melyet én akkor még fejből is tudtam – így fest:

x	0	1	2	3	4	5	6	7	8	9	...
$p(x)$	0.06	0.17	0.24	0.22	0.16	0.09	0.04	0.02	0.01	0.00	...

(Lásd a "Poisson eloszlás $\lambda = 2.81$ paraméterrel" című ábrát is!)

A táblázatból is és az ábrából is jól látható (amit a Poisson eloszlás móduszára vonatkozó ismereteinkből a táblázat és az ábra nélkül is tudunk), hogy az eloszlás módusza 2 -vel egyenlő, vagyis 2 hal a legvalószínűbb. Ezért 2 halra tippeltem – ez volt a legjobb tipp.

Természetesen a 3 hal, vagy az 1 hal, vagy a 4 hal sem rossz tipp, bár ezek a tippek kicsit kisebb valószínűséggel vezetnek sikerre. Viszont – például – a 9, 10, 11 vagy több hal valószínűségei – mint számolással ellenőrizhető – még együttesen is 0.001 -nél kevesebbet tesznek ki, ezért az ilyen tippek nem kecsegtetnek a siker reményével.

2. feladat: Hány hal lesz az *IFJÚ* halász hálójában? Érdekes utána járni, hogy hogyan kellett volna gondolkodnom, ha a halász fiatal lett volna, és – sok éves tapasztalat hiányában – így felelt volna:

"A legutóbbi 100 merítés során kifigyeltem, hogy pontosan 6 -szor volt üres a háló!"

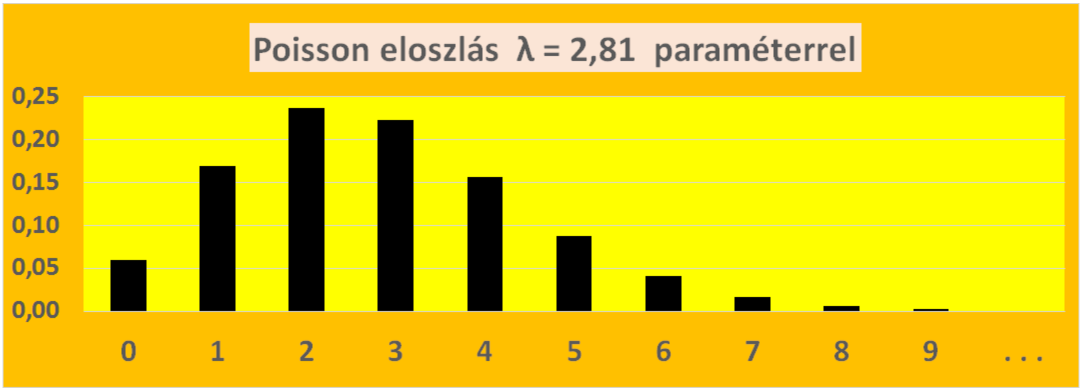
Igaz, hogy 6 osztva 100 -zal, megfelel az előző történetben mondott 6 % -nak, de mégis mást jelent. Az öreg halász által mondott 6 % -ot – mivel sok éves tapasztalatra épült – a 0 darab hal valószínűségének tekinthettük. Az ifjú halász által mondott "100 merítés során 6 -szor volt üres a háló" kijelentés csak egyetlen (100 merítésből álló) megfigyelés!

Abból a tényből, hogy egy alkalommal 100 merítésből 6 -szor üres a háló, nem vehetjük a 0 darab hal valószínűségét 6 % -nak. Az, hogy 100 merítésből 6 -szor üres a háló, megtörténhet még akkor is, ha a 0 darab hal valószínűsége eltér 0.06 -től.

A 0 darab hal ismeretlen valószínűségét jelöljük p -vel. A p segítségével könnyen felírhatjuk annak valószínűségét, hogy 100 merítésből pontosan 6-szor üres a háló:

$$\text{P(100 merítésből 6 -szor üres a háló)} = \binom{100}{6} p^6 (1-p)^{100-6}$$

A jobb oldalon álló képlet p függvényeként egyszerűen vizsgálható:



2. ábra. *Poisson eloszlás* $\lambda = 2.81$ *paraméterrel*

p	**0.015**	0.030	0.045	**0.060**	0.075	0.090	0.105	0.120	0.135	**0.150**
$\binom{100}{6}p^6(1-p)^{100-6}$	**0.003**	0.050	0.131	**0.166**	0.139	0.089	0.047	0.022	0.009	**0.003**

A táblázatból is kitűnik, és az analízis eszközeivel is könnyen belátható, hogy a

$$\binom{100}{6}p^6(1-p)^{100-6}$$

kifejezés értéke

- $p < 0.060$ esetén növekszik
- $p = 0.060$ esetén a maximális érték 0.166 -dal egyenlő
- $0.060 < p$ esetén csökken

A táblázatban vastagítással jelezzük a tényt, hogy $p =$**0.015** -nél is és $p =$**0.150** -nél is a kifejezés értéke csak **0.003**. A 0.003 érték a maximális 0.166 értékhez képest elhanyagolhatóan kicsi. Ezért a p -vel jelölt ismeretlen valószínűséget 0.015 és 0.150 közöttinek feltételezzük. Leginkább a $p = 0.060$ érték tűnik jogosnak, de a 0.015 és 0.150 közötti p értékeket sem zárjuk ki. A $p < 0.015$ és a $p > 0.150$ lehetőségeket viszont már elvetjük.

Ez a döntés kétségtelenül egy szubjektív döntés, melyhez hasonlókat a mindennapi életben nap mint nap megteszünk: például szívesbben ülünk egy olyan autóba, amiben van légzsák, mint egy olyanba, amelyikben nincs. A személyi sérülés valószínűsége egy esetleges balesetnél a légzsák esetén lényegesen kisebb, mintha nem lenne légzsák.

Az előző feladat megoldásában elmagyaráztuk, hogy a hálóban lévő halak száma – mint valószínűségi változó – Poisson eloszlást követ. Az eloszlás paraméterét a

$$\frac{\lambda^0}{0!}\mathrm{e}^{-\lambda} = p \qquad \text{azaz} \qquad \mathrm{e}^{-\lambda} = p$$

egyenletből a $\lambda = -\ln(p)$ formula alapján kapjuk meg:

p	0.015	0.030	0.045	0.060	0.075	0.090	0.105	0.120	0.135	0.150
$\lambda = -\ln(p)$	**4.20**	3.51	3.10	2.81	2.59	2.41	2.25	2.12	2.00	**1.90**

E táblázatban vastagítással kiemeltük a $p = 0.015$ és a $p = 0.150$ szélső esetekhez tartozó λ értékeket:

- $p = 0.015$ esetén $\lambda = \mathbf{4.20}$
- $p = 0.150$ esetén $\lambda = \mathbf{1.90}$

Ezekhez a paraméter értékekhez tartozó Poisson eloszlások táblázatai így festenek:

Poisson eloszlás $\lambda = \mathbf{4.20}$ paraméterrel

x	0	1	2	3	**4**	5	6	7	8	9	...
$p(x)$	0.015	0.063	0.132	0.185	**0.194**	0.163	0.114	0.069	0.036	0.017	...

Poisson eloszlás $\lambda = \mathbf{1.90}$ paraméterrel

x	0	**1**	2	3	4	5	6	7	8	9	...
$p(x)$	0.150	**0.285**	0.270	0.171	0.081	0.031	0.010	0.003	0.001	0.000	...

A táblázatokból látható (amit a Poisson eloszlás móduszára vonatkozó ismereteinkből a táblázatok nélkül is tudunk):

- $\lambda = 4.20$ (vagyis $p = 0.015$) esetén a Poisson eloszlás legvalószínűbb tagja, azaz módusza 4 -gyel egyenlő
- $\lambda = 1.90$ (vagyis $p = 0.150$) esetén a Poisson eloszlás legvalószínűbb tagja, azaz módusza 1 -gyel egyenlő

A $0.015 < p < 0.150$ valószínűségeknek megfelelő $1.90 < \lambda < 4.20$ paraméter értékek melletti Poisson eloszlások móduszai nyilván 1 és 4 közé esnek.

Ezért – összegezve a gondolatokat – ezt kapjuk: *lehetséges, hogy akár* 1 *vagy* 2 *vagy* 3 *vagy* 4 *hal a legvalószínűbb, de a* 2 *hal tűnik a legjobb tippnek*, hiszen – fentebb – láttuk, hogy a

$$\mathrm{P}(\text{ 100 merítésből 6 -szor üres a háló}) = \binom{100}{6} p^6 (1-p)^{100-6}$$

valószínűség $p = 0.060$ esetén a legnagyobb, ami $\lambda = 2.81$ -nek felel meg, és $\lambda = 2.81$ esetén a Poisson eloszlás legvalószínűbb tagja, vagyis a módusza 2 -vel egyenlő.

Ez az eredmény összhangban van az előző feladat megoldásával, de ugyanakkor érzékelteti, hogy nem mindegy, hogy valaki a sok éves tapasztalata alapján képes egy valószínűséget megmondani, vagy csak 100 kísérlet eredménye alapján a relatív gyakorisággal közelíti azt.

3. feladat: Kullancsok a futóversenyen. A "Kocogj velünk!" mozgalom keretében tavaly futóversenyt rendeztek a Duna-kanyarban. A pályát sajnos kullanccsal fertőzött területen át vezették. Kiderült, hogy a versenyzők közül 300-an találtak magukban egy, 75-en pedig két kullancsot. Ennek alapján becsüljük meg, hogy körülbelül hányan indultak a versenyen!

Megoldás első része:

Ha azt vizsgáljuk, hogy egy kiszemelt versenyzőben, mondjuk a "Futó Botond" nevűben, hány kullancs lesz a verseny után, akkor egy valószínűségi változót kapunk. Mivel a sok kullancs mindegyike – a többitől függetlenül – kis valószínűséggel kerül ebbe a versenyzőbe, ez a valószínűségi változó Poisson eloszlást követ valamilyen λ praméterrel. Ezért – a Poisson eloszlás képlete szerint – annak a valószínűsége, hogy 1 kullancs kerül Futó Botondba, $\lambda e^{-\lambda}$. A 2 kullancs valószínűsége pedig $\frac{\lambda^2}{2} e^{-\lambda}$. Ha a versenyzők számát N -nel jelöljük, és a valószínűségeket relatív gyakoriságokkal helyettesítjük, akkor az alábbi egyenleteket állíthatjuk fel:

$$\frac{300}{N} = \lambda e^{-\lambda}$$

$$\frac{75}{N} = \frac{\lambda^2}{2} e^{-\lambda}$$

Ezt az egyenletrendszert könnyű megoldani. Először elosztjuk a második egyenletet az elsővel. A kapott új egyenletben a baloldal törtben N -nel, a jobboldaliban λ -val és $e^{-\lambda}$ -val egyszerűsítve ezt kapjuk:

$$\frac{75}{300} = \frac{\lambda}{2}$$

vagyis

$$\lambda = \frac{150}{300} = 0.5$$

Ezek után az első egyenletből N -re ez jön ki:

$$N = \frac{300}{\lambda e^{-\lambda}} = \frac{300}{0.5\ e^{-0.5}} = 989.2$$

A versenyzők száma természetesen egész szám. Az, hogy itt N -re nem egész jött ki, annak a következménye, hogy a kiszemelt versenyzőben az 1, illetve 2 kullancs valószínűségét relatív gyakoriságokkal közelítettük. Ezért a feladatban feltett kérdésre kézenfekvő a közelítő válasz: **Körülbelül 1000 versenyző volt a versenyen**.

Megoldás második része - *(Extra tananyag)*:

Az emberben óhatatlanul felmerül a kérdés: vajon mennyire körülbelül a "körülbelül 1000"? A probléma elemzése érdekében kiszámoljuk most, hogy mi a valószínűsége annak, hogy 1000 versenyző esetén

pontosan 300 versenyzőben lesz 1 kullancs

Ha mindenegyes versenyzővel kapcsolatban az 1 kullancs valószínűségét $\frac{300}{1000} = 0.3$ -nak vesszük, és feltételezzük, hogy a versenyzőkben a kullancsok száma egymástól független, akkor a keresett valószínűségre az 1000 -ed rendű, 0.3 paraméterű binomiális eloszlás szerint Excellel a

```
BINOM.DIST( 300 ; 1000 ; 0.3 ; FALSE )
```

képlet adódik. A valószínűség numerikus értéke 0.027. Azt is kiszámolhatjuk, hogy mi a valószínűsége annak, hogy 1000 versenyző esetén

pontosan 75 versenyzőben lesz 2 kullancs

Ha mindenegyes versenyzővel kapcsolatban a 2 kullancs valószínűségét $\frac{75}{1000} = 0.075$ -nek vesszük, és feltételezzük, hogy a versenyzőkben a kullancsok száma egymástól független, akkor a keresett valószínűségre az 1000 -ed rendű, 0.075 paraméterű binomiális eloszlás szerint Excellel a

```
BINOM.DIST( 75 ; 1000 ; 0.075 ; FALSE )
```

képlet adódik. A valószínűség numerikus értéke 0.048. Azt is kiszámoljuk, hogy mi a valószínűsége annak, hogy 1000 versenyző esetén

pontosan 300 versenyzőben lesz 1 kullancs, és
pontosan 75 versenyzőben lesz 2 kullancs

azaz

pontosan 300 versenyzőben lesz 1 kullancs, és
pontosan 75 versenyzőben lesz 2 kullancs, és
1000-300-75 versenyzőben lesz 0 kullancs vagy 2-nél több kullancs

Ha – úgyanúgy, mint fentebb – mindenegyes versenyzővel kapcsolatban az 1 kullancs valószínűségét $\frac{300}{1000} = 0.3$ -nak, a 2 kullancs valószínűségét $\frac{75}{1000} = 0.075$ -nek vesszük, és feltételezzük, hogy a versenyzőkben a kullancsok száma egymástól független, akkor a keresett valószínűség az 1000 -ed rendű, $(0.3; 0.075)$ paraméterű polinomiális eloszlással adódik:

$$\frac{(1000)!}{(300)!\,(75)!\,(1000-300-75)!}\;(0.3)^{(300)}\,(0.075)^{(75)}\,(1-0.3-0.075)^{(1000-300-75)}$$

A polinomiális eloszlás nincs beépítve az Excelbe, és a polinomiális eloszlás képletében szereplő faktoriálisokat sem tudja az Excel kiszámolni. Viszont a polinomiális eloszlást kétdimenziós normális eloszlással közelíthetjük, aminek eredményeképpen a valószínűségre (hat tizedesjegyre kerekítve) 0.001269 jön ki. Azon, hogy egy nagyon kis valószínűség érték jött ki, nem szabad meglepődni: érezhetően kicsi annak esélye, hogy 1000 versenyző közül

pontosan 300 versenyzőben lesz 1 kullancs, és
pontosan 75 versenyzőben lesz 2 kullancs

Az embernek eszébe jut, hogy elvileg olyan szélsőséges helyzet is lehetne, hogy – mondjuk – csak 375 versenyző indul a versenyen, és a sors úgy hozza, hogy

300 versenyzőben lesz 1 kullancs, és
75 versenyzőben lesz 2 kullancs

Avagy 5000 versenyző esetén sem zárható ki, hogy

300 versenyzőben lesz 1 kullancs, és
75 versenyzőben lesz 2 kullancs
4625 versenyzőben pedig 0 kullancs vagy 2-nél több kullancs

Ezért elbizonytalanodunk, hogy az egyenletrendszerből adódó 989, avagy 1000, amit kerekítéssel kaptunk, vajon tényleg jó közelítése a versenyzők számának? A bizonytalanság eloszlatása céljából elemezzük most a problémát, és megnézzük, hogy különböző N és λ értékek esetén mi a valószínűsége annak, hogy

pontosan 300 versenyzőben lesz 1 kullancs, és
pontosan 75 versenyzőben lesz 2 kullancs

A valószínűségeket az Excel segítségével számoltunk ki, és táblázatba rendezve adjuk meg. Mivel a valószínűségek értéke nagyon pici, a táblázatban a valószínűségek értékeinek millószorosát (egészekre kerekítve) írtuk be:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	λ
750	0	0	0	0	0	0	10	1	0	
800	0	0	0	0	0	35	39	0	0	
850	0	0	0	0	2	374	10	0	0	
900	0	0	0	0	96	388	2	0	0	
950	0	0	0	0	833	55	2	0	0	
1000	0	0	0	1	1 294	1	0	0	0	
1050	0	0	0	16	448	0	0	0	0	
1100	0	0	0	125	41	0	0	0	0	
1150	0	0	0	295	1	0	0	0	0	
1200	0	0	0	242	0	0	0	0	0	
1250	0	0	0	77	0	0	0	0	0	
N										

A táblázatból kitűnik, hogy – arányaikat tekintve – "kicsi" és "kicsi" között is nagy a különbség! Annak a valószínűsége, hogy

pontosan 300 versenyzőben lesz 1 kullancs, és
pontosan 75 versenyzőben lesz 2 kullancs

a táblázatban látható értékek közül 1000 versenyző esetén a legnagyobb, és már 950 vagy 1050 versenyző esetén is jóval kisebb, de 900 vagy annál kevesebb, illetve 1100 vagy annál több versenyző esetén sokkal-sokkal kisebb.

Ezért józan, elfogadható következtetésnek tűnik: körülbelül 1000 versenyző indult a versenyen, ahol a "körülbelül" azt jelenti, hogy 950-nél kevesebb vagy 1050-nál több versenyző gyakorlatilag kizárt.

9.5.2. Poisson eloszlás kétdimenzióban

Súlyfüggvény:

$$p(x,y) = \frac{\lambda_1^x}{x!}\, e^{-\lambda_1}\ \frac{\lambda_2^y}{y!}\, e^{-\lambda_2} \qquad \text{ha} \quad x,y = 0,1,2,\ldots$$

Mikor használjuk: Amikor egy olyan kétdimenziós valószínűségi változóval van dolgunk, melynek koordinátái függetlenek, és külön-külön Poisson eloszlást követnek valamilyen paraméterekkel.

Paraméterek jelentése:

$\lambda_1 =$ az első koordináta "átlagos" értéke (pontosabban: várható értéke)

$\lambda_2 =$ a második koordináta "átlagos" értéke (pontosabban: várható értéke)

A várható érték fogalmát a következő fejezetben tanuljuk.

9.6. Gyakorló feladatok

1. A tanult nevezetes eloszlások közül melyik illik legjobban az alábbi valószínűségi változók modellezésére?
   - Nevezze meg eloszlást!
   - Indokolja meg az eloszlás jogosságát!
   - Ha lehet, adja meg az eloszlás paramétereinek numerikus értékét!
   - Ha nem lehet, magyarázza el, milyen adatokra, mérési eredményekre lenne szükség a paraméterek megadásához!

   Íme a valószínűségi változók:

   (a) Egy érmét 15-ször feldobunk, és megnézzük, hogy hányszor kapunk fejet.
   (b) Egy érmét 15-ször feldobunk, és megnézzük, hogy hányszor kapunk írást.
   (c) Két érmét 15-ször feldobunk, és megnézzük, hogy hányszor kapunk dupla fejet.
   (d) Két érmét 15-ször feldobunk, és megnézzük, hogy hányszor kapunk legalább egy fejet.
   (e) Egy dobókockát 15-ször feldobunk, és megnézzük, hogy hányszor kapunk hatost.
   (f) Addig dobunk egy szabályos érmével, amíg végre fejet kapunk, és megfigyeljük, hány dobás kell ehhez.
   (g) Addig dobunk egy szabályos érmével, amíg másodszorra fejet kapunk, és megfigyeljük, hány dobás kell ehhez.

(h) Addig dobunk két szabályos érmével, amíg harmadszorra kapunk dupla fejet, és megfigyeljük, hány dobás kell ehhez.

(i) Addig dobunk egy szabályos dobókockával, amíg végre hatost kapunk, és megfigyeljük, hogy hány dobás kellett ehhez.

(j) Addig dobunk egy szabályos dobókockával, amíg másodszorra hatost kapunk, és megfigyeljük, hogy hány dobás kellett ehhez.

(k) Addig dobunk egy szabályos dobókockával, amíg harmadszorra hatost kapunk, és megfigyeljük, hogy előtte hányszor dobtunk hattól különböző számot.

(l) Ahányadik autó felvesz fel, amikor kiállok az országútra (mert autóstoppal akarok utazni).

(m) Ahány autó elmegy 10 perc alatt.

(n) Ahányadik autó a harmadik piros.

(o) Ahány pótkocsi nélküli elmegy az első pótkocsis előtt.

(p) Ahány pótkocsi nélküli elmegy az ötödik pótkocsis előtt.

(q) Ahány kocsi megáll az első 10 odaérkező kocsi közül.

(r) Ahány kocsi jön 5 perc alatt.

(s) Ahány kocsi jön 10 perc alatt.

(t) Ahány kocsi megáll 10 perc alatt.

(u) Ahány kocsi elmegy, mielőtt az első megáll.

(v) Ahányadik az a kocsi, amelyik másodiknak megáll.

(w) Ahány kocsi elmegy, mielőtt a második megáll.

(x) Ahány kocsi elmegy, mielőtt a harmadik megáll.

2. Feltéve, hogy egy 100 tagú társaságban 40 balkezes van, adjuk meg annak a valószínűségét, hogy 10 véletlenszerűen kiválasztott ember között

   (a) pontosan 3 balkezes van.

   (b) pontosan x balkezes van.

3. Feltéve, hogy egy országban a balkezesek aránya kb. 40%, adjuk meg annak a valószínűségét, hogy 10 véletlenszerűen kiválasztott ember között

   (a) pontosan 3 balkezes van.

   (b) pontosan x balkezes van.

4. Egy gyárban futószalag szállítja az alkatrészeket. A futószalag leáll, ha selejtes termék érkezik. A termékek 2%-a selejtes. Mi az eloszlása annak a valószínűségi változónak, ami azt számolja, hogy

   (a) hányszor állt le a szalag az n -edik termékig (őt is beleértve)?

   (b) hány terméket gyártott a gép az n -edik leállásig?

   (c) hány terméket szállított két leállás között?

   (d) hány leállás történt egymás után anélkül, hogy egyetlen jó termék is keletkezett volna?

5. Egy forgalmas országútszakaszon, ahol máskor is szoktak radarozni, figyelik, hogy 5 perc alatt hány autó lépi át a megengedett sebességhatárt. Tapasztalat szerint kb. ugyanolyan valószínű, hogy lesz ilyen autó, mint az, hogy nem lesz. Mennyi a valószínűsége, hogy az 5 perc alatt pontosan három autó lépi át a megengedett sebességhatárt?

6. Egy osztályban 22 tanuló van. Egy órára 8-an nem készültek, és 7-en felelnek. Adjuk meg a készületlen felelők számának eloszlását! Mennyi a valószínűsége, hogy pontosan 2 készületlen felelő lesz?

7. 100 kulcs közül 2 nyitja az előttünk lévő ajtót. A kipróbált kulcsokat félretesszük. Mi a valószínűsége, hogy legfeljebb 50 próbálkozásból bejutunk? És mi a valószínűsége, hogy pontosan n próbálkozásból jutunk be?

8. Tegyük fel, hogy hétköznaponként Pesten 0.4, Budán 0.7 annak a valószínűsége, hogy nem történik súlyos baleset. Mi a valószínásége annak, hogy

   (a) Pesten pontosan 1 baleset történik?

   (b) Budán pontosan 1 baleset történik?

   (c) Budapesten pontosan 1 baleset történik?

10. Módusz megkeresése

10.1. Előkészületek *(Extra tananyag)*

Könnyű megérteni az alábbi három triviális tényt, ha az Olvasó papírt és ceruzát ragad, és – például – pálcikákkal szemlélteti a szóbanforgó eloszlásokat.

1. tény: Ha egy eloszlás $p(k)$ súlyfüggvényére igaz, hogy

- $p(k-1) < p(k)$ ha $k \leq k_0$ (a súlyfüggvény k_0 -ig növekszik)
- $p(k-1) > p(k)$ ha $k > k_0$ (utána csökken)

akkor ez azt jelenti, hogy

- a $p(0), p(1), \ldots, p(k_0)$ sorozat szigorúan monoton növekedő
- a $p(k_0), p(k_0+1), \ldots$ sorozat szigorúan monoton csökkenő

ezért a súlyfüggvény a maximális értékét a k_0 helyen veszi fel,
vagyis az eloszlás módusza a k_0 szám.

2. tény: Ha egy eloszlás $p(k)$ súlyfüggvényére igaz, hogy

- $p(k-1) < p(k)$ ha $k < k_0$ (a súlyfüggvény $(k_0 - 1)$ -ig növekszik)
- $p(k-1) = p(k)$ ha $k = k_0$ ($(k_0 - 1)$ -nél és k_0 -nál egyforma az értéke)
- $p(k-1) > p(k)$ ha $k > k_0$ (utána csökken)

akkor ez azt jelenti, hogy

- a $p(0), p(1), \ldots, p(k_0 - 1)$ sorozat szigorúan monoton növekedő
- $p(k_0 - 1) = p(k_0)$
- a $p(k_0), p(k_0+1), \ldots$ sorozat szigorúan monoton csökkenő

ezért a súlyfüggvény a maximális értékét a $(k_0 - 1)$ és a k_0 helyeken veszi fel,
vagyis az eloszlásnak két módusza van: $(k_0 - 1)$ és k_0 .

3. tény: Ha $\lfloor x_0 \rfloor$ -lal jelöljük az x_0 valós szám egész részét, vagyis azt az egész számot, amit x_0 -ból kapunk, ha lefelé kerekíjük, és k egész szám, akkor

- a $k \leq x_0$ egyenlőtlenség ekvivalens a $k \leq \lfloor x_0 \rfloor$ egyenlőtlenséggel
- a $k > x_0$ egyenlőtlenség ekvivalens a $k > \lfloor x_0 \rfloor$ egyenlőtlenséggel

E három tény együtt adja a következő módszert.

10.2. Módszer a módusz képletének meghatározására

Egyetlen módusz esete: Ha egy eloszlás $p(k)$ súlyfüggvényére és egy x_0 valós számra igaz, hogy

- $p(k-1) < p(k)$ minden $k \leq x_0$ -ra, és
- $p(k-1) > p(k)$ minden $k > x_0$ -ra,

akkor az eloszlás módusza az $\lfloor x_0 \rfloor$ szám.

Két módusz esete: Ha egy eloszlás $p(k)$ súlyfüggvényére és egy x_0 egész számra igaz, hogy

- $p(k-1) < p(k)$ ha $k < x_0$ -ra
- $p(k-1) = p(k)$ ha $k = x_0$ -ra
- $p(k-1) > p(k)$ ha $k > x_0$ -ra

akkor az eloszlásnak két módusza van: $x_0 - 1$ és x_0 .

1. Példa: Binomiális eloszlás módusza. A binomiális eloszlás súlyfüggvénye

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, 2, \ldots)$$

Ezért a $p(k-1) < p(k)$ egyenlőtlenség most ezt jelenti:

$$\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} < \binom{n}{k} p^k (1-p)^{n-k}$$

Triviális egyszerűsítések és alakítások következnek:

$$\frac{n!}{(k-1)!(n-k+1)!} \, p^{k-1} (1-p)^{n-k+1} \; < \; \frac{n!}{k!(n-k)!} \, p^k (1-p)^{n-k}$$

$$\frac{1-p}{n-k+1} \; < \; \frac{p}{k}$$

$$k\,(1-p) \; < \; (n-k+1)\,p$$

$$k \; - \; k\,p \; < \; n\,p \; - \; k\,p \; + \; p$$

$$k \; < \; n\,p \; + \; p$$

Tehát a $p(k-1) < p(k)$ egyenlőtlenség ekvivalens a

$$k < (n+1)\,p$$

egyenlőtlenséggel. A $p(k-1) > p(k)$ egyenlőtlenség pedig nyilván ekvivalens a

$$k > (n+1)\,p$$

egyenlőséggel. Ez azt jelenti, hogy $(n+1)\,p$ tölti be azt a szerepet, amit a módusz meghatározására mondott módszerben az x_0 töltött be. Ezért ha $(n+1)\,p$ nem egész szám, akkor az eloszlásnak egyetlen módusza a $\lfloor (n+1)\,p \rfloor$ szám. Ha $(n+1)\,p$ egész szám, $p(k-1) = p(k)$ is fenn tud állni $k = (n+1)\,p$ esetén. Ezért ha $(n+1)\,p$ egész szám, akkor az eloszlásnak két módusza van: $(n+1)\,p - 1$ és $(n+1)\,p$.

2. Példa: Poisson eloszlás módusza. A Poisson eloszlás súlyfüggvénye

$$p(k) = \frac{\lambda^k}{k!}\, e^{-\lambda} \qquad (k = 0, 1, 2, \ldots)$$

Ezért a $p(k-1) < p(k)$ egyenlőtlenség most ezt jelenti:

$$\frac{\lambda^{k-1}}{(k-1)!}\, e^{-\lambda} < \frac{\lambda^k}{k!}\, e^{-\lambda}$$

Triviális egyszerűsítések adják, hogy az egyenlőtlenség ekvivalens ezzel:

$$k < \lambda$$

A $p(k-1) > p(k)$ egyenlőtlenség pedig ekvivalens a

$$k > \lambda$$

egyenlőséggel. Ez azt jelenti, hogy λ tölti be azt a szerepet, amit a módusz meghatározására mondott módszerben az x_0 töltött be. Ezért ha λ nem egész szám, akkor az eloszlásnak egyetlen módusza a $\lfloor \lambda \rfloor$ szám. Ha λ egész szám, akkor $p(k-1) = p(k)$ is fenn tud állni $k = \lambda$ esetén. Ezért ha λ egész szám, akkor az eloszlásnak két módusza van: $\lambda - 1$ és λ .

10.3. Nevezetes eloszlások móduszai – formulák

Hipergeometrikus eloszlás

Ha $(n+1)\,\frac{M+1}{N+2}$ egész szám, akkor két módusz van: $(n+1)\,\frac{M+1}{N+2} - 1$ és $(n+1)\,\frac{M+1}{N+2}$

Ha $(n+1)\,\frac{M+1}{N+2}$ nem egész szám, akkor egy módusz van: $\left\lfloor (n+1)\,\frac{M+1}{N+2} \right\rfloor$

Binomiális eloszlás

Ha $(n+1)\,p$ egész, akkor két módusz van: $(n+1)\,p - 1$ és $(n+1)\,p$

Ha $(n+1)\,p$ nem egész, akkor egy módusz van: $\lfloor (n+1)\,p \rfloor$

Indikátor eloszlás

Ha $p = \frac{1}{2}$, akkor két módusz van: 0 és 1

Ha $p < \frac{1}{2}$, akkor egy módusz van: 0

Ha $p > \frac{1}{2}$, akkor egy módusz van: 1

Optimista geometriai eloszlás

módusz: 1

Pesszimista geometriai eloszlás

módusz: 0

Optimista negatív binomiális eloszlás

Ha $\frac{r-1}{p}$ egész szám, akkor két módusz van: $\frac{r-1}{p}$ és $\frac{r-1}{p}+1$

Ha $\frac{r-1}{p}$ nem egész szám, akkor egy módusz van: $\left\lfloor\frac{r-1}{p}\right\rfloor+1$

Pesszimista negatív binomiális eloszlás

Ha $\frac{r-1}{p}$ egész, akkor két módusz van: $\frac{r-1}{p}-r$ és $\frac{r-1}{p}-r+1$

Ha $\frac{r-1}{p}$ nem egész, akkor egy módusz van: $\left\lfloor\frac{r-1}{p}\right\rfloor-r+1$

Poisson eloszlás

Ha λ egész, akkor két módusz van: $\lambda - 1$ és λ

Ha λ nem egész, akkor egy módusz van: $\lfloor\lambda\rfloor$

10.4. Gyakorló feladatok

1. Vezesse le a hipergeometrikus eloszlás móduszára fentebb megadott képletet!
2. Vezesse le a negatív binomiális eloszlás móduszára fentebb megadott képletet!
3. Tegyük fel, hogy valaki azzal szórakozik, hogy mindenkivel, akivel összefut, megkérdezi, hogy mikor van a születésnapja. (Az év nem számit, csak a hónap, nap. Szökőévektől eltekintünk. Feltételezzük, hogy minden megkérdezett ember a többitől függetlenül egyforma eséllyel mondja születésnapjaként az év 365 napját.) Emberünk izgul, hogy hányadik embernél adódik az első olyan helyzet, hogy a megkérdezett ember születésnapja egybeesik egy olyan születésnappal, amit már valaki mondott korábban: vajon hányadok kérdezésnél következik ez be? Határozza meg ennek a valószínűségi változónak
   (a) a lehetséges értékeit!
   (b) súlyfüggvényét!
   (c) móduszát!
4. Tegyük fel, hogy egy X valószínűségi változó binomiális eloszlást követ, melynek paramétereit nem ismerjük. Azt azért tudjuk, hogy az n paraméter 10 és 15 között van, és azt is tudjuk, hogy X legvalószínűbb értéke 5. Adjon alsó és felső becslést az $X = 6$ esemény valószínűségére!
5. Sok éves tapasztalatok alapján tudni lehet, hogy egy bizonyos országban az a legvalószínűbb, hogy egy év alatt 3 hármasiker születik. Adjon alsó és felső becslést annak az eseménynek a valószínűségére, hogy egy évben nem születnek hármasikrek!

11. Szimuláció

11.1. A [0;1] intervallum felosztásának módszere

Vegyünk fel 1-nél kisebb pozitív számoknak egy véges sok tagból álló, növekedő sorozatát! Íme – például – egy 3 tagból sorozat:

$$0.2\,, \quad 0.6\,, \quad 0.7$$

Most egy táblázatot alkotunk, melynek első sora 0 -val kezdődik, aztán következnek a sorozat tagjai, a második sorba pedig írjunk négy akármilyen számot, például a 10, 20, 30, 40 számokat:

0	0.2	0.6	0.7
10	20	30	40

Ha egy ilyen táblázatot készítünk Excelben, akkor

angolul: `HLOOKUP( RAND() ; * ; 2 )` magyarul: `VKERES( VÉL() ; * ; 2 )`

utasítás, ahol * helyén a táblázatra való hivatkozás áll, számunkra előnyösen működik. A generált véletlen számot összehasonlítja a táblázat első sorában lévő számokkal, és megkeresi a sorban azt az elemet, amelyik egyenlő vagy éppen alatta van a generált véletlen értéknek, és az utasítás értékéül a második sorból veszi azt az elemet, amit az első sorban talált elem alatt van. Tehát

- ha a véletlen érték 0 és 0.2 közé esik, vagy 0 -val egyenlő, akkor a függvény értéke 10
- ha a véletlen érték 0.2 és 0.6 közé esik, vagy 0.2 -del egyenlő, akkor a függvény értéke 20
- ha a véletlen érték 0.6 és 0.7 közé esik, vagy 0.6 -del egyenlő, akkor a függvény értéke 30
- ha a véletlen érték 0.7 és 1 közé esik, vagy 0.7 -del egyenlő, akkor a függvény értéke 40

Mivel

- a $[\,0\,;\;0.2\,]$ intervallum hossza 0.2
- a $[\,0.2\,;\;0.6\,]$ intervallum hossza 0.4
- a $[\,0.6\,;\;0.7\,]$ intervallum hossza 0.1
- a $[\,0.7\,;\;1\,]$ intervallum hossza 0.3

a függvény értéke

- 0.2 valószínűséggel 10
- 0.4 valószínűséggel 20
- 0.1 valószínűséggel 30
- 0.3 valószínűséggel 40

Tehát a

`HLOOKUP( RAND() ; * ; 2 )` `VKERES( VÉL() ; * ; 2 )`

Excel függvénnyel generált valósszínűségi változó eloszlása

x	10	20	30	40
$p(x)$	0.2	0.4	0.1	0.3

Könnyű észrevenni, hogy milyen számokat kell az Excel táblázatba írni ahhoz, hogy adott véges diszkrét eloszlást követő valószínűségi változót kapjunk ezzel a módszerrel.

Például a megadott eloszlás legyen a következő:

x	1	2	3	4	5	6
$p(x)$	0.05	0.10	0.15	0.3	0.25	0.15

Ekkor az Excel táblázatot így kell elkészíteni:

0	0.05	0.15	0.30	0.60	0.85
1	2	3	4	5	6

Az Excel táblázat első sorában álló számokat a 0 után összegzéssel kaptuk meg a megadott eloszlás valószínűségeiből:

$$0.05 = 0 + 0.05$$

$$0.15 = 0 + 0.05 + 0.10$$

$$0.30 = 0 + 0.05 + 0.10 + 0.15$$

$$0.60 = 0 + 0.05 + 0.10 + 0.15 + 0.30$$

$$0.85 = 0 + 0.05 + 0.10 + 0.15 + 0.30 + 0.25$$

11.2. Gyakorló feladatok

1. Szimuláljon Excellel olyan valószínűségi változót, ami binomiális eloszlást követ
   (a) $n = 5$ és $p = 0.4$ paraméterekkel.
   (b) $n = 5$ és változtatható p paraméterekkel.
   (c) változtatható $n = 5$ és p paraméterekkel.
2. Szimuláljon Excellel olyan valószínűségi változót, ami optimista geometriai eloszlást követ változtatható p paraméterrel.
   (a) $p = 1/2$ paraméterrel.
   (b) $p = 1/6$ paraméterrel.
   (c) változtatható p paraméterrel.
3. Szimuláljon Excellel olyan valószínűségi változót, ami az alábbi táblázattal megadot eloszlást követi:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

4. Szimuláljon Excellel olyan valószínűségi változót, ami az alábbi táblázattal megadot eloszlást követi:

x	k_0	k_1	k_2	k_3
$p(x)$	0.2	0.4	0.3	0.1

ahol k_0, k_1, k_2, k_3 változtatható paraméterek.

5. Szimuláljon Excellel olyan valószínűségi változót, ami az alábbi táblázattal megadot eloszlást követi:

x	k_0	k_1	k_2	k_3
$p(x)$	p_0	p_1	p_2	p_3

ahol k_0, k_1, k_2, k_3 és p_0, p_1, p_2, p_3 változtatható paraméterek.

12. Tömegpont rendszerek

Előrebocsátunk két, a fizikából ismert képletet. Tekintsünk egy tömegpont-rendszert a számegyenesen, mely egy véges vagy megszámlálhatóan végtelen sok pontból álló S halmazra koncentrálódik. Ha x eleme S -nek, akkor az x pontban lévő tömeg mennyiségét jelöljük $p(x)$ -szel. A pontrendszer **súlypont**ja – mint ismeretes

$$\frac{\sum\limits_x x\; p(x)}{\sum\limits_x p(x)}$$

A kifejezésben itt – és a későbbiekben is – a szummázás az összes lehetséges x -re értendő. A

$$\sum_x (x-c)^2\; p(x)$$

érték pedig a pontrendszernek **a** c **pontra vonatkozó tehetetlenségi nyomatéka**.

Megjegyzés: Vegyük észre, hogy ha az össztömeg 1-gyel egyenlő, azaz $\sum\limits_x p(x) = 1$, akkor a súlypontot megadó képlet így egyszerűsödik:

$$\sum_x x\; p(x)$$

13. Egydimenziós adatrendszerek

13.1. Átlag

Adatoknak (számoknak) egy sorozatát *adatrendszer*nek hívjuk. Egy adatrendszer valamilyen értelemben vett közepét, az *átlag* fogalmát, mindenki ismeri.

1. Példa: Adatrendszer 5 elemből. A könnyebb követhetőség kedvéért az alábbi adatrendszer csak 5 elemből áll:

adatok	54	55	44	44	60

A sor végén feltüntetjük az átlagot, ami 51.4:

						átlag
adatok	54	55	44	44	60	51.4

Az adatok átlagát *első momentum*nak is szokás hívni. Egy külön sorban feltüntetjük az adatok négyzeteit is és azok átlagát is:

						átlag
adatok	54	55	44	44	60	51.4
adatok négyzetei	2916	3025	1936	1936	3600	2 682.6

13.2. Második momentuma

Az adatok négyzeteinek az átlagát az adatrendszer *második momentum*ának (esetleg 0 *-ra vonatkozó második momentum*ának) hívjuk. A példánkban szereplő adatrendszer második momentuma $2\,682.6$. Az összegét – például Excellel – mátrixok szorzataként is ki lehet számolni. Az alábbi mátrix szorzás, majd osztás az adatok számával (ami itt 5) a második momentumot adja:

$$\boxed{\begin{matrix} 54 & 55 & 44 & 44 & 60 \end{matrix}} \cdot \boxed{\begin{matrix} 54 \\ 55 \\ 44 \\ 44 \\ 50 \end{matrix}} \cdot \frac{1}{5} = 2\,682.6$$

Ha a négyzetre emelés előtt egy c számot kivonunk az adatrendszer minden tagjából, akkor a c *-re vonatkozó második momentum*ot kapjuk meg. Az alábbi táblázatban $c = 50$, és az 50 -re vonatkozó második momentum 42.6:

						átlag
adatok	54	55	44	44	60	
adatok és c különbségei	4	5	-6	-6	10	
különbségek négyzete	16	25	36	361	100	42.6

A $c = 50$ -re vonatkozó (itt $c = 50$) második momentumot – a (0-ra vonatkozó) "síma" második momentumhoz hasonlóan – a megfelelő sor- és oszlopvektorok szorzásával, és utána a darabszámmal való osztással is megkaphatjuk:

$$\boxed{\begin{array}{ccccc} 4 & 5 & -6 & -6 & 10 \end{array}} \cdot \boxed{\begin{array}{c} 4 \\ 5 \\ -6 \\ -6 \\ 10 \end{array}} \cdot \frac{1}{5} = 42.6$$

A $c = 50$ -re vonatkozó második momentum arról ad felvilágosítást, hogy

- az adatrendszer vajon a c érték köré tömörül-e – ilyenkor a $c = 50$ -re vonatkozó második momentum értéke kisebb , vagy
- a $c = 50$ értéktől jobban szóródik – ilyenkor a c -re vonatkozó második momentum értéke nagyobb

13.3. Variancia, szórás

Speciálisan kiszámolhatjuk az adatoknak az átlagtól, itt most 51.4 -től vett különbségét is:

						átlag
adatok	54	55	44	44	60	51.4
adatok és átlag különbségei	2.6	3.6	-7.4	-7.4	8.6	0

A kapott adatrendszert az eredeti adatrendszer *centralizált*jának hívjuk. A centralizált adatok azt mutatják, hogy az eredeti adatok hogyan helyezkednek el az átlagukhoz viszonyítva. Ha a centralizált érték pozitív, akkor az eredeti adat az átlag felett van, ha ha a centralizált érték negatív, akkor az eredeti adat az átlag alatt van. A centralizált érték nagysága pedig az átlagtól való távolsággal egyenlő.

Egy centralizált adatrendszer átlaga nyilván mindig 0. A centralizált adatrendszer második momentumát az eredeti adatrendszer *varianciá*jának vagy *szórásnégyzeté*nek hívjuk. A mi adatrendszerünk varianciája 40.64. A variancia négyzetgyökét, aminek numerikus értéke itt 6.37, *szórás*nak nevezzük.

						átlag	
centralizált adatok	2.6	3.6	-7.4	-7.4	8.6	0	
centralizált adatok négyzetei	6.76	12.96	54.76	54.76	73.96	40.64	6.37
						variancia	szórás

Hangsúlyozzuk, hogy a varianciát – például Excellel – mátrixok szorzásával is ki lehet számolni. A centralizált adatokból alkotott sorvektort jobbról szorozva a transzponáltjával, majd osztva az adatok számával (ami itt 5), a varianciát kapjuk:

$$\boxed{\begin{array}{ccccc} 2.6 & 3.6 & -7.4 & -7.4 & 8.6 \end{array}} \cdot \boxed{\begin{array}{c} 2.6 \\ 3.6 \\ -7.4 \\ -7.4 \\ 8.6 \end{array}} \cdot \frac{1}{5} = 40.64$$

A variancia és a szórás arról adnak felvilágosítást, hogy

- az adatrendszer az átlag köré összetömörül-e – ilyenkor a variancia értéke kisebb, vagy
- jobban szóródik – ilyenkor a variancia értéke nagyobb

A következő két adatrendszer mindgyikének az átlaga 100, de a variancia és a szórás a második adatrendszer esetében nagyobb, mint az első esetében Ezzel tudjuk numerikusan kifejezni azt a szemléletes tényt, hogy a második adatrendszer jobban szóródik az átlag körül, mint az első.

adatok	99	100	97	99	105	9	3
						variancia	szórás

adatok	95	100	85	95	125	225	15
						variancia	szórás

Megjegyzés: Elég természetes dolognak tűnne, hogy egy adatrendszer szóródását az átlagtól való (mindenféle négyzetre emeléstől mentes) távolságoknak az átlagával, az ún. *átlagos abszolút eltérés*sel jellemezzük:

						átlag	
adatok	54	55	44	44	60	51.4	
adatok és átlag különbségei	2.6	3.6	-7.4	-7.4	8.6	0	(mindig 0)
távolságok az átlagtól	2.6	3.6	7.4	7.4	8.6	5.92	átlagos abszolút eltérés

Ezzel a természetes fogalommal szépen lehet numerikusan számolni, dolgozni, de – mint az később kiderül – a variancia és a szórás olyan technikai előnyökkel bírnak, amelyek mind az elméletben, mind a gyakorlatban háttérbe szorítják az átlagos abszolút eltérést. Megjegyezzük, hogy egy adatrendszer átlagos abszolút eltérése legfeljebb akkora lehet, mint a szórása.

Vegyük észre, hogy a kiindulásul felvett adatrendszer varianciáját úgy is megkaphatjuk, hogy a második momentumból kivonjuk az első momentum négyzetét:

$$40.64 = 2\,682.6 - (51.4)^2$$

Ez az összefüggés mindig igaz:

$$\text{variancia} = \text{második momentum} - (\text{ első momentum })^2$$

A bizonyítást az Olvasóra bízzuk.

Megjegyzés: Minden olyan kalkulátor, ami tud statisztikai üzemmódban dolgozni, egy-egy gombnyomásra kiadja a betátplált adatrendszer átlagát, varianciáját, szórását. Az Excelben az átlagra, varianciára, szórásra beépített függvények vannak.

Egyszerű algebrai összefüggés a Steiner egyenlőség, és annak következménye a Steiner egyenlőtlenség:

Steiner egyenlőség: Akármilyen c -re igaz, hogy

$$c \text{ -re vonatkozó második momentum} = \text{variancia} + (c \text{ - átlag })^2$$

Steiner egyenlőtlenség: Akármilyen c -re igaz, hogy

$$c \text{ -re vonatkozó második momentum} \geq \text{variancia}$$

$c = \text{ átlag }$ esetén az egyenlőség, más c értékekre az egyenlőtlenség teljesül.

13.4. Medián

Itt teszünk említést az a*datrendszer mediánjá*nak fogalmáról.

- Ha az adatokat növekvő sorrendbe rendezzük, akkor – páratlan sok tag esetén – a sorrendben egyértelmű középső tagot nevezzük mediánnak.
-
  - Ha páros sok tag van, akkor két "szinte középső" tag van. Vannak, akik ezeknek az átlagát nevezik mediánoknak
  - Mások mindkettőt mediánnak nevezik.
  - És vannak akik nemcsak ezt a két értéket, hanem minden közöttük lévő számot is mediánnak neveznek.

  Páros sok tag esetén három kissé különböző deinició jól elfér egymás mellett és a fejünkben is. Mindhárom hozzállásnak van előnye is, hátránya is.

  - Az első definíció szerint a medián egyértelmű.
  - A második definíció esetén mindkét medián az adathalmaz eleme. Ha valamiért kell, akkor az átlagukat bármikor ki lehet számítani.
  - A harmadik definíció a legbátrabb és sok szempontból a legegészségesebb: a mediánok egy zárt intervallumot alkotnak. Az intervallum belsejében minden elemre vitathatatlanul igaz, hogy felezi az adathalmazt: tőla balra és jobbra az adathalmaznak ugyanannyi eleme van.

  Páros sok tag esetén az Excelben a `MEDIAN` függvény az első definíciónak megfelelően működik.

14. Valószínűségi változók és eloszlások várható értéke, varianciája, szórása

14.1. Várható érték

Tekintsünk egy diszkrét X valószínűségi változót. X lehetséges értékeinek halmazát jelöljük S -sel, az S valamely elemét x -szel, az x -hez tartozó valószínűséget pedig $p(x)$ -szel. Ekkor a

$$\sum_x x\, p(x)$$

számot az X valószínűségi változó **várható érték**ének nevezzük (Várható érték angolul: expected value). A várható értéket $\mathrm{E}(X)$ -szel jelöljük:

$$\mathrm{E}(X) = \sum_x x\, p(x)$$

A kifejezésben – mint korábban is és a későbbiekben is – a szummázás az összes lehetséges x -re értendő.

Ha egy eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk, akkor a

$$\sum_x x\, p(x)$$

számot az **eloszlás várható érték**ének nevezzük.

1. Megjegyzés: Ezeket a képleteket a súlypont képletével összevetve azonnal szembetűnik, hogy $\sum_x p(x) = 1$ feltétel teljesülése miatt a súlypont képlete a várható érték képletébe megy át. Tehát egy valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a várható érték egybeesik a tömegeloszlás súlypontjával.

Kicsit általánosabban, ha a szóbanforgó valószínűségi változó mellett még egy $y = t(x)$ transzformációt is tekintünk, és képezzük a $t(X)$ valószínűségi változót, akkor a

$$\mathrm{E}\,(\,t(X)\,) = \sum_x t(x)\, p(x)$$

számot a transzformációval nyert $t(X)$ **valószínűségi változó várható érték**ének nevezzük.

Ha egy eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk, akkor a

$$\sum_x t(x)\, p(x)$$

számot a szóbanforgó eloszlásból **a t transzformációval kapott eloszlás várható értéké**nek nevezzük.

Még általánosabban, egy (X, Y) kétdimenziós valószínűségi változó esetén ha még egy $z = t(x, y)$ transzformációt is tekintünk, és képezzük a $t(X, Y)$ valószínűségi változót, akkor a

$$\mathrm{E}\,(\,t(X,Y)\,) = \sum_{(x,y)} t(x,y)\, p(x,y)$$

számot a transzformációval nyert $t(X, Y)$ **valószínűségi változó várható érték**ének nevezzük.

Ha egy síkbeli eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk, akkor a

$$\sum_{(x,y)} t(x,y)\, p(x,y)$$

számot a szóbanforgó eloszlásból **a t transzformációval kapott eloszlás várható értéké**nek nevezzük.

2. Megjegyzés: Ha az $y = t(x)$ függvény a négyzetre emelés függvénye, azaz $t(x) = x^2$, akkor a

$$\sum_x x^2 \, p(x)$$

kifejezéshet jutunk. Ezt az értékét a valószínűségi változó avagy a szóbanforgó eloszlás **második momentum**ának nevezzük. A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a második momentum egybeesik a tömegeloszlásnak a 0 pontra vonatkozó tehetetlenségi nyomatékával.

3. Megjegyzés: Ha a t függvény az identitás függvény, azaz $t(x) = x$, akkor a

$$\sum_x x \, p(x)$$

kifejezéshet jutunk, ami a várható érték képlete. Ezt a képletet a második momentummal összevetve látjuk, hogy a várható értéket jogos **első momentum**nak is nevezni.

4. Megjegyzés: Ha a lehetséges értékeket indexezve soroluk fel: $x_1, x_2, \ldots$, és a hozzájuk tartozó valószínűségeket pedig $p_1, p_2, \ldots$ -vel jelöljük, akkor a képletek így festenek:

$$\mathrm{E}(X) = \sum_i x_i \, p_i$$

$$\mathrm{E}\,(\,t(X)\,) = \sum_i t(x_i) \, p_i$$

5. Megjegyzés: Jelöljük X -szel a kockadobás eredményét. X^2 várható értékét szemléletesen úgy is kiszámolhatnánk, hogy a dobókocka oldalaira új címkéket ragasztanánk $1, 4, 9, 16, 25, 36$ számokkal. Így nem meglepő, hogy $\mathrm{E}(t(X))$ képletében a $p(x)$ értékek változatlanok maradnak, míg x -eket mindenhol formálisan kicseréljük $t(x)$ -re.

6. Megjegyzés: Ha X lehetséges értékeinek S halmaza végtelen sok elemet tartatlmaz, akkor a fenti sorok végtelen sorok, ezért ilyenkor elofordulhat, hogy a sor nem konvergens, hanem értéke plusz- vagy mínusz végtelen, vagy akár az is, hogy a sor értéke nem definált, mert a sor pozitív tagjaiból álló sor és a negatív tagjaiból álló sor is divergál, és ezért a $-\infty + \infty$ alakú, nem értelmezhető eset áll elő. Ha X lehetséges értékeinek S halmaza csak véges sok elemet tartalmaz, akkor ilyen nem fordulhat elő, és a várható érték egy jól definiált szám.

14.2. Feltételes várható érték egy eseményen belül

Példa: Fiatal házaspár egy vagy több gyerekkel. Korábbi példánk szerint, ha egy fiatal házaspártházaspárt véletlenszerűen választunk, és tekintjük az

$$X = \text{ gyerekek száma}$$

valószínűségi változót, akkor a valószínűségi változó eloszlása:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

Egyszerű számolással kapjuk, hogy az X valószínűségi változó várható értéke

$$\mathrm{E}(X) = 0 \cdot 0.2 + 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.1 = 1.3$$

Tehát ha véletlenszerűen választunk sok családot, akkor körülbelül 1.3.

Kérdés: Ha a kiválasztott családok közül csak a gyerekes családok gyerekei számának az átlagát vesszük, akkor ez az átlag körülbelül mennyivel egyenlő?

A válasz egyszerű: korábban meghatároztuk az X valószínűségi változó $X \geq 1$ feltétel melletti (feltételen belüli) feltételes eloszlását:

x	1	2	3
$p(x)$	$\frac{4}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Ennek várható értéke:

$$1 \cdot \frac{4}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{13}{8} = 1.625$$

Általánosan is megfogalmazzuk a példában leírtakat: Ha S a számegyenesnek egy részhalmaza, és tekintünk egy

$$\{p(x),\ x \in S\}$$

eloszlást S -en, továbbá A részhalmaza S -nek, akkor a

$$\left\{ \frac{p(x)}{P(A)},\ x \in A \right\}$$

feltételes eloszlás

$$\sum_{x \in A} x \ \frac{p(x)}{P(A)}$$

várható értékét az eredeti eloszlásnak **az A esemény bekövetkezése melletti feltételes várható értéké**nek vagy **az A -n belüli feltételes várható értéké**nek nevezzük, és $\mathrm{E}(X \mid A)$ -val jelöljük.

A fenti példával kapcsolatban ezt írhatjuk:

$$\mathrm{E}(X \mid X \geq 1) = 1.625$$

14.3. Geometriai eloszlás jellemzése feltételes várható értékekkel *(Extra tananyag)*

Az alábbi két állítás mindegyike fontos jellemzése a geometriai eloszlásoknak. Az elsőnek a bizonyítása triviális, nem is adjuk meg. A másodiknak a bizonyítása kissé nehéznek tűnhet, de igazából egyszerű és érdekes.

1. Állítás: Ha egy pozitív egész értékű, diszkrét X valószínűségi változóra teljesül, hogy az

$$\mathrm{P}(\, X = s+1 \,|\, X \;>\; s\,)$$

feltételes valószínűség nem függ s -től, akkor X optimista geometriai eloszlást követ.

Bizonyítás: Triviális.

2. Állítás: Ha egy pozitív értékű, diszkrét X valószínűségi változóra teljesül, hogy az

$$\mathrm{E}(\, X \;-\; s \,|\, X \;>\; s\,)$$

feltételes várható érték nem függ s -től, akkor X optimista geometriai eloszlást követ.

Bizonyítás: Az $\mathrm{E}(X - s|X > s)$ feltételes várható értéket felírjuk összegekkel:

$$\mathrm{E}(\, X - s \,|\, X \;>\; s\,) \;=\; \frac{1\,p_{s+1} + 2\,p_{s+2} + 3\,p_{s+3} + \ldots}{p_{s+1} + p_{s+2} + p_{s+3} + \ldots}$$

Ha ez a feltételes várható érték egy c konstanssal egyenlő, akkor minden s -re, igaz, hogy

$$\frac{1\,p_{s+1} + 2\,p_{s+2} + 3\,p_{s+3} + \ldots}{p_{s+1} + p_{s+2} + p_{s+3} + \ldots} = c$$

A nevezővel átszorozva ez adódik:

$$1\,p_{s+1} + 2\,p_{s+2} + 3\,p_{s+3} + \ldots \;\;=\; c\,(p_{s+1} + p_{s+2} + p_{s+3} + \ldots)$$

Írjuk fel ezt a legutolsó egyenletet s helyett $s-1$ -re is:

$$1\,p_{s} + 2\,p_{s+1} + 3\,p_{s+2} + \ldots \;\;=\; c\,(p_{s} + p_{s+1} + p_{s+2} + \ldots)$$

Az alsó egyenletből kivonva a felsőt, a következő, minden s -re fennálló egyenletet kapjuk:

$$p_s + p_{s+1} + p_{s+2} + \ldots \;=\; c\,p_s$$

Írjuk fel ugyanezt az egyenletet s helyett $s+1$ -re:

$$p_{s+1} + p_{s+2} + p_{s+3} + \ldots \;=\; c\,p_{s+1}$$

Most vonjuk ki a felsőből az alsót, ezt kapjuk:

$$p_s \;=\; c\;(\,p_s - p_{s+1}\,)$$

ahonnan

$$c\,p_{s+1} \;=\; (c-1)\;p(s)$$

vagyis

$$p_{s+1} \;=\; \frac{c-1}{c}\,p(s)$$

ami világosan mutatja, hogy a $p_1, p_2, p_3, \ldots$ sorozat minden eleme az előző elemből egy konstanssal való szorzással adódik, ezért a sorozat egy geometriai sorozatot alkot. Tehát tényleg egy geometriai eloszlásról van szó.

Geometriai eloszlás egy számsorozat elemein: Ha az $a_1, a_2, \ldots$ sorozat egy számtani sorozatot alkot, akkor

$$p(a_k) = (1-p)^{k-1}p \qquad (k = 1, 2, \ldots)$$

képlettel definiált eloszlás neve: **geometriai eloszlás a** $\{a_1, a_2, \ldots\}$ **halmazon**.

14.4. Variancia és szórás

Ha c egy szám, akkor a valószínűségi változó, illetve az eloszlás c -re vonatkozó második momuentumának nevezzük a

$$\sum_x (x-c)^2 \; p(x)$$

értéket. A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a c -re vonatkozó második momuentum egybeesik a tömegeloszlásnak a c pontra vonatkozó tehetetlenségi nyomatékával. Ha c az eloszlás várható értéke, akkor a második momuentum neve **variancia** vagy **szórásnégyzet**, amit $\text{VAR}(X)$ -szel jelölünk:

$$\text{VAR}(X) = \sum_x (x - \text{E}(X))^2 \; p(x)$$

A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a variancia egybeesik a tömegeloszlásnak a súlypontjára vonatkozó tehetetlenségi nyomatékával.

A variancia négyzetgyökét, a

$$\sqrt{\sum_x (x-c)^2 \; p(x)}$$

számot **szórás**nak hívjuk, és $\text{SD}(X)$ -vel jelöljük:

$$\text{SD}(X) = \sqrt{\text{VAR}(X)} = \sqrt{\sum_x (x - \text{E}(X))^2 \; p(x)}$$

1. Megjegyzés: Emlékeztetünk rá, hogy egy valószínűségi változó, illetve egy eloszlás 0 -ra vonatkozó második momentumát, a

$$\sum_x x^2 \; p(x)$$

számot egyszerűen csak **második momuentum**nak nevezzük. Könnyű belátni, hogy igaz a

$$\text{variancia} = \text{második momentum} - (\text{ első momentum })^2$$

összefüggés.

2. Megjegyzés: Nyilvánvalóan igaz a Steiner egyenlőség, és annak következménye a Steiner egyenlőtlenség:

Steiner egyenlőség: Akármilyen c -re igaz, hogy

$$\text{E}\left((X-c)^2\right) = \text{VAR}(X) + (\; c \; - \text{E}(X)\;)^2$$

Steiner egyenlőtlenség: Akármilyen c -re igaz, hogy

$$\text{E}\left((X-c)^2\right) \geq \text{VAR}(X)$$

$c \; = \text{E}(X)$ esetén az egyenlőség, más c értékekre az egyenlőtlenség teljesül.

14.5. Gyakorló feladatok

1. Tegyük fel, hogy egy normált tömeg eloszlás az $[A; B]$ intervallumra koncentrálódik. Hogy néz ki az eloszlás, ha
   (a) a súlypontja a lehető leginkább balra helyzkedik el?
   (b) a súlypontja a lehető leginkább jobbra helyzkedik el?
   (c) ha a tehetetlenségi nyomatéka a lehető legnagyobb?

2. A

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

   táblázattal megadot eloszlásnak mennyi a
   (a) várható értéke?
   (b) második momentuma?
   (c) varianciája?
   (d) szórása?
   (e) a 2.5 pontra vonatkozó második momentuma?

3. A

x	10	11	12	13
$p(x)$	0.2	0.4	0.3	0.1

   táblázattal megadot eloszlásnak mennyi a
   (a) várható értéke?
   (b) második momentuma?
   (c) varianciája?
   (d) szórása?
   (e) a 2.5 pontra vonatkozó második momentuma?

4. A

x	0	10	20	30
$p(x)$	0.2	0.4	0.3	0.1

táblázattal megadot eloszlásnak mennyi a

(a) várható értéke?
(b) második momentuma?
(c) varianciája?
(d) szórása?
(e) a 2.5 pontra vonatkozó második momentuma?

5. A

x	0	11	22	33
$p(x)$	0.2	0.4	0.3	0.1

táblázattal megadot eloszlásnak mennyi a

(a) várható értéke?
(b) második momentuma?
(c) varianciája?
(d) szórása?
(e) a 2.5 pontra vonatkozó második momentuma?

6. Tegyük fel, hogy egy bizonyos országban a családoknak kb.

- 15%-ának nincs gyereke
- 40%-ának 1 gyereke van
- 30%-ának 2 gyereke van
- 10%-ának 3 gyereke van
- 5%-ának pedig 4

A 4 -nél többgyerekes családok olyan ritkák, hogy ezzel a lehetőséggel nem foglalkozunk. Feltesszük, hogy a különböző családokban a gyerekek száma független egymástól. Feltesszük, hogy minden gyerek a többitől függetlenül $0.5 - 0.5$ valószínűségggel születik lánynak vagy fiúnak.

(a) Átlagosan hány gyerek van egy-egy családban?
(b) Átlagosan hány lány-gyerek van egy-egy családban?
(c) Átlagosan hány fiú-gyerek van egy-egy családban?
(d) Ha egy családról annyit tudunk, hogy van benne gyerek, akkor a gyerekek számának mi az eloszlása?
(e) A családi pótlékot az alábbi táblázat szerint kapják a családok:

gyerekek száma	0	1	2	3	4
családi pótlék (*fitying*-ben)	$0\ 000$	$5\ 000$	$25\ 000$	$30\ 000$	$35\ 000$

Átlagosan hány fitying családi pótlékot kap egy-egy család?

(f) Átlagosan hány fitying családi pótlékot kapnak a gyerekes családok?

7. Egy dobozban pénzérmék vannak. Az érmék súlyai:

érme	**5 Ft**	**10 Ft**	**20 Ft**	**50 Ft**	**100 Ft**	**200 Ft**
tömeg (g)	4.2	6.1	7.0	7.7	8.0	9.5

Feltételezzük, hogy húzáskor az érmék valószínűségei arányosak a súlyaikkal. 1000 -szer húzunk visszatevéssel. Kb. mennyi lesz a kihúzott érmék

- értkének
- súlyának

az átlaga? (Ha esetleg kísérletekkel ellenőrizni akarja számítását, akkor azt ne igazi érmékkel, hanem számítógépes szimulációval tegye! Az, hogy húzáskor az érmék valószínűségei arányosak a súlyaikkal, csak egy hipozézis volt. Egyáltalán nem biztos, hogy ezek a valószínűségek megfelelnek a valóságnak.)

8. Találja meg az általános képletet egy valószínűségi változó várható értékére, ha az eloszlását nem a súlyfüggvénnyel, hanem a lehetséges értékek súlyaival adjuk meg (melyek összege általában nem 1.)

15. Nagy számok törvényei

Mind elméleti, mind gyakorlati szempontból fontos tény, hogy nagy számú kísérletek eredményeinek átlaga, szórása és sok egyéb jellemzője – bár függnek a véletlentől – mégis egy fajta stabilitást mutatnak: értékük közel van egy olyan értékhez, amit elméleti úton – a kísérletek elvégzése nélkül – is meg lehet határozni. Erre a tényre, mint a nagy számok törvényeire szoktunk hivatkozni.

15.1. NSZT a kísérleti eredmények átlagára

Heurisztikus megfogalmazás: Képzeljük el, hogy egy X valószínűségi változóra sok kísérletet végzünk. A kísérletek számát jelölje N, a kísérleti eredményeket, melyek véletlen számok, jelölje $X_1, X_2, \ldots, X_N$. Ha kísérleti eredményeket átlagoljuk, akkor ez az átlag – nagy kísérletszám esetén – körülbelül az X várható értékével egyenlő:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \mathrm{E}(\, X \,)$$

Emlékeztetőül:

$$\mathrm{E}(\, X \,) = \sum_x x\, p(x)$$

Vázlatos bizonyítás: Az X valószínűségi változó lehetséges értékei legyenek

$$x_1,\ x_2, \ldots$$

Jelölje

$$N_1,\ N_2, \ldots$$

a lehetséges értékek gyakoriságait az N kísérlet során. Tehát minden i esetén N_i azt mutatja, hogy x_i hányszor következik be az N kísérlet során. Tudjuk, hogy minden i -re az N_i/N relatív gyakoriság – nagy N esetén – közel van a p_i valószínűséghez. Az

$$\frac{X_1 + X_2 + \ldots + X_N}{N}$$

kifejezés számlálója N darab tagból áll. Közöttük az x_1 érték N_1 -szer, az x_2 érték N_2 -ször, ... fordul elő. Ezért a számláló értéke

$$X_1 + X_2 + \ldots + X_N = N_1 \cdot x_1 + N_2 \cdot x_2 + \ldots$$

a tört értéke pedig

$$\frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{N_1 \cdot x_1 + N_2 \cdot x_2 + \ldots}{N} =$$

$$= \frac{N_1}{N} \cdot x_1 + \frac{N_2}{N} \cdot x_2 + \ldots \approx p_1 \cdot x_1 + p_2 \cdot x_2 + \ldots \ = \ \mathrm{E}(\, X \,)$$

Precíz megfogalmazás *(Extra tananyag)*:

1. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\, X \,) = \sum_x x\, p(x)$ sor értéke egyértelműen definiált véges szám, ami teljesül, ha

   - a sor véges sok tagból áll, azaz az X valószínűségi változónak csak véges sok lehetséges értéke van, vagy
   - végtelen sok tag esetén – ha a sor abszolút konvergens, azaz $\sum_x |x|\ p(x) < \infty$.

   Ekkor $N \to \infty$ esetén az

   $$\frac{X_1 + X_2 + \ldots + X_N}{N} \ \to \ \mathrm{E}(\, X \,)$$

   esemény valószínűsége 1-gyel egyenlő.

2. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\ X\ ) = \sum_x x\ p(x)$ sor értéke egyértelműen plusz végtelen, ami akkor teljesül, ha

   (a) pozitív tagokat tartalmazó $\sum_{x>0}\ x\ p(x)$ divergens, és

   (b) a negatív tagokat tartalmazó $\sum_{x<0}\ x\ p(x)$ konvergens.

   Ekkor $N \to \infty$ esetén az
   $$\frac{X_1 + X_2 + \ldots + X_N}{N}\ \to\ +\infty$$
   esemény valószínűsége 1-gyel egyenlő.

3. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\ X\ ) = \sum_x x\ p(x)$ sor értéke egyértelműen mínusz végtelen, ami akkor teljesül, ha

   (a) pozitív tagokat tartalmazó $\sum_{x>0}\ x\ p(x)$ konvergens, és

   (b) a negatív tagokat tartalmazó $\sum_{x<0}\ x\ p(x)$ divergens.

   Ekkor $N \to \infty$ esetén az
   $$\frac{X_1 + X_2 + \ldots + X_N}{N}\ \to\ -\infty$$
   esemény valószínűsége 1-gyel egyenlő.

15.2. NSZT a kísérleti eredmények függvényének az átlagára

Heurisztikus megfogalmazás: Képzeljük el, hogy egy X valószínűségi változóra sok kísérletet végzünk. A kísérletek számát jelölje N, a kísérleti eredményeket, melyek véletlen számok, jelölje $X_1, X_2, \ldots, X_N$. Ha kísérleti eredményeket egy $t(x)$ függvénybe helyettesíthetjük, majd az így kapott
$$t(X_1)\ ,\ t(X_2)\ ,\ \ldots, t(X_N)$$
értékeket átlagoljuk, akkor ez az átlag – nagy kísérletszám esetén – körülbelül a $t(X)$ várható értékével egyenlő:
$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N}\ \approx\ \mathrm{E}\left(\ t(X)\ \right)$$
Emlékeztetőül:
$$\mathrm{E}\left(\ t(X)\ \right) = \sum_x t(x)\ p(x)$$

Vázlatos bizonyítás: Az X valószínűségi változó lehetséges értékei legyenek
$$x_1,\ x_2, \ldots$$
Jelölje
$$N_1,\ N_2, \ldots$$
a lehetséges értékek gyakoriságait az N kísérlet során. Tehát minden i esetén N_i azt mutatja, hogy x_i hányszor következik be az N kísérlet során. Tudjuk, hogy minden i -re az N_i/N relatív gyakoriság – nagy N esetén – közel van a p_i valószínűséghez. A
$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N}$$

kifejezés számlálója N darab tagból áll. Közöttük a t függvény argumentumaként az x_1 érték N_1 -szer, az x_2 érték N_2 -ször, ... fordul elő. Ezért a számláló értéke

$$t(X_1) + t(X_2) + \ldots + t(X_N) = N_1 \cdot t(x_1) + N_2 \cdot t(x_2) + \ldots$$

a tört értéke pedig

$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} = \frac{N_1 \cdot t(x_1) + N_2 \cdot t(x_2) + \ldots}{N} =$$

$$= \frac{N_1}{N} \cdot t(x_1) + \frac{N_2}{N} \cdot t(x_2) + \ldots \approx p_1 \cdot t(x_1) + p_2 \cdot t(x_2) + \ldots \;=\; \mathrm{E}(\, t(X)\,)$$

Precíz megfogalmazás *(Extra tananyag)*:

1. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\, t(X)\,) = \sum_x t(x)\; p(x)$ sor értéke egyértelműen definiált véges szám, ami teljesül, ha

   (a) a sor véges sok tagból áll, azaz az X valószínűségi változónak csak véges sok lehetséges értéke van, vagy

   (b) végtelen sok tag esetén – ha a sor abszolút konvergens, azaz $\sum_x |t(x)|\;\; p(x) < \infty$.

   Ekkor $N \to \infty$ esetén a
   $$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \;\to\; \mathrm{E}(\, t(X)\,)$$
   esemény valószínűsége 1-gyel egyenlő.

2. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\, t(X)\,) = \sum_x t(x)\; p(x)$ sor értéke egyértelműen plusz végtelen, ami akkor teljesül, ha

   (a) pozitív tagokat tartalmazó $\sum_{t(x)>0}\; x\; p(x)$ divergens, és

   (b) a negatív tagokat tartalmazó $\sum_{t(x)<0}\; x\; p(x)$ konvergens.

   Ekkor $N \to \infty$ esetén a
   $$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \;\to\; +\infty$$
   esemény valószínűsége 1-gyel egyenlő.

3. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\, t(X)\,) = \sum_i t(x_i)\; p_i = \sum_i t(x)\; p(x)$ sor értéke egyértelműen mínusz végtelen, ami akkor teljesül, ha

   (a) pozitív tagokat tartalmazó $\sum_{t(x)>0}\; x\; p(x)$ kovergens, és

   (b) a negatív tagokat tartalmazó $\sum_{t(x)<0}\; x\; p(x)$ divergens.

   Ekkor $N \to \infty$ esetén a
   $$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \;\to\; -\infty$$
   esemény valószínűsége 1-gyel egyenlő.

15.3. NSZT a második momentumra

Az $X_1, X_2, \ldots, X_N$ kísérleti eredmények második momentumára igaz, hogy nagy kísérletszám esetén körülbelül az X második momentumával egyenlő:

$$\frac{X_1^2 + X_2^2 + \ldots + X_N^2}{N} \approx \mathrm{E}(\, X^2 \,)$$

Emlékeztetőül:

$$\mathrm{E}(\, X^2 \,) = \sum_x x^2 \, p(x)$$

Bizonyítás: Ha az előző részben a $t(x) = x^2$ speciális függvényt vesszük akkor azonnal megkapjuk az állítást.

15.4. NSZT a varianciára

Az $X_1, X_2, \ldots, X_N$ kísérleti eredmények

$$\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_N - \overline{X})^2}{N}$$

varianciájára igaz, hogy nagy kísérletszám esetén körülbelül az X varianciájával egyenlő:

$$\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_N - \overline{X})^2}{N} \approx \mathrm{VAR}(X)$$

Emlékeztetőül:

$$\mathrm{VAR}(X) = \sum_x (x - \mathrm{E}(X))^2 \, p(x)$$

ahol

$$\mathrm{E}(\, X \,) = \sum_x x \, p(x)$$

Vázlatos bizonyítás: Ugyanis, ha az $\overline{X}_N$ átlag helyére az $\mathrm{E}(\, X \,)$ várható értéket írjuk, akkor azt kapjuk, hogy

$$\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_N - \overline{X})^2}{N} \approx$$

$$\approx \frac{(X_1 - \mathrm{E}(\, X \,))^2 + (X_2 - \mathrm{E}(\, X \,))^2 + \ldots + (X_N - \mathrm{E}(\, X \,))^2}{N}$$

A jobb oldalon álló átlag az $(X - \mathrm{E}(\, X \,))^2$ valószínűségi változó várható értékéhez van közel, ami éppen az X varianciájával egyenlő:

$$\frac{(X_1 - \mathrm{E}(\, X \,))^2 + (X_2 - \mathrm{E}(\, X \,))^2 + \ldots + (X_N - \mathrm{E}(\, X \,))^2}{N} \approx \mathrm{E}\left((X - \mathrm{E}(\, X \,))^2\right) = \mathrm{VAR}(X)$$

a

15.5. NSZT a szórásra

Így aztán nagy kísérletszám esetén az $X_1, X_2, \ldots, X_N$ kísérleti eredmények szórása – igen általános feltételek mellett – körülbelül az X szórásával egyenlő:

$$\sqrt{\frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\ldots+(X_N-\overline{X})^2}{N}} \approx \mathrm{SD}(X)$$

Emlékeztetőül:

$$\mathrm{SD}(X) = \sqrt{\sum_x (x-\mathrm{E}(X))^2\, p(x)}$$

ahol

$$\mathrm{E}(\,X\,) = \sum_x x\, p(x)$$

15.6. NSZT a mediánra

Mivel az adatrendszerek mediánját és az eloszlások mediánját is szavakkal írtuk le, így a mediánra vonatkozó NSZT -ét is csak szavakkal adju meg:

Az $X_1, X_2, \ldots, X_N$ kísérleti eredmények mediánjára igaz, hogy nagy kísérletszám esetén körülbelül az X mediánjával egyenlő.

15.7. Gyakorló feladatok

1. Egy dobozban 6 cédula van, rajtuk pedig a következő számok:

   (a) $1,\ 2,\ 3,\ 4,\ 5,\ 6;$

   (b) $1,\ 2,\ 6,\ 6,\ 6,\ 6;$

   (c) $1,\ 2,\ 2,\ 3,\ 3,\ 3;$

   (d) $1,\ 2,\ 2,\ 2,\ 2,\ 6;$

   (e) $1,\ 1,\ 1,\ 2,\ 2,\ 3;$

   (f) $11,\ 11,\ 11,\ 12,\ 12,\ 13;$

   (g) $21,\ 21,\ 21,\ 22,\ 22,\ 23;$

   (h) $21,\ 22,\ 22,\ 22,\ 22,\ 26;$

   (i) $210,\ 220,\ 220,\ 220,\ 220,\ 236;$

   (j) $0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6;$

   (k) $0.1,\ 0.2,\ 0.6,\ 0.6,\ 0.6,\ 0.6;$

   (l) $5.1,\ 5.2,\ 5.6,\ 5.6,\ 5.6,\ 5.6;$

   (m) $-1,\ -2,\ -3,\ -4,\ -5,\ -6;$

   (n) $-1,\ -2,\ -6,\ -6,\ -6,\ -6;$

   (o) $-1,\ -2,\ 2,\ 3,\ 3,\ 3.$

A fenti esetek mindegyikében véletlenszerűen húzunk a dobozból egy cédulát, és leolvassuk a rajta lévő számot. Ezt a számot jelöljük X -szel. Képzelje el (még jobb ha ténylegesen vagy szimulációval meg is teszi), hogy X -re sok kísérletet végez. A kísérletek számát jelölje N, a kísérleti eredményeket, amelyek véletlen számok, jelölje

$$X_1, X_2, \ldots, X_N$$

- Becsülje meg,hogy körülbelül mennyi lesz a kísérleti eredmények négyzetének átlaga, vagyis az

  $$\frac{X_1^2 + X_2^2 + \ldots + X_N^2}{N}$$

  kifejezés értéke? (Az itt szereplő kifejezést az $X_1, X_2, \ldots, X_N$ számok második momentumának nevezzük. Az elméleti értéket, amit Önnek kell meghatározni, az X valószínűségi változó (avagy a valószínűségi változó eloszlása) második momentumának nevezzük.)
- Becsülje meg, hogy körülbelül mennyi lesz a kísérleti eredményeknek egy valamilyen konstanstól, például a 2.8-től való eltérése négyzetének az átlaga, vagyis az

  $$\frac{(X_1 - 2.8)^2 + (X_2 - 2.8)^2 + \ldots + (X_N - 2.8)^2}{N}$$

  kifejezésnek az értéke? (Az itt szereplő kifejezést az $X_1, X_2, \ldots, X_N$ számoknak a 2.8 -re vonatkozó második momentumának nevezzük. Az elméleti értéket, amit Önnek meg kell határozni, az X valószínűségi változó (avagy a valószínűségi változó eloszlása) 2.8 -re vonatkozó második momentumának nevezzük.)
- Becsülje meg,hogy körülbelül mennyi lesz a kísérleti eredményeknek az átlagukra vonatkozó második momentuma. Az elméleti értéket, amit Önnek meg kell határozni, az X valószínűségi változónak a várható értékre vonatkozó második momentumának nevezzük. Ezt a mennyiséget varianciának, avagy szórásnégyzetnek is hívjuk.

2. Először számolja ki a

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

eloszlás

(a) várható értékét
(b) második momentumát
(c) varianciáját
(d) szórását

majd pedig szimuláljon olyan valószínűségi változót, amely a megadott eloszlást követi, csináljon sok kísérletet, számolja ki a kísérleti eredmények

(a) átlagát
(b) második momentumát
(c) varianciáját
(d) szórását

és győződjön meg arról, hogy a kísérleti eredmények megfelelő jellemzői az eloszlás megfelelő jellemzőihez közel vannak, a kísérletsorozat ismételgetésével azok körül ingadoznak.

16. Várható érték, variancia, szórás általános tulajdonságai

16.1. Várható érték tulajdonságai

A várható érték, a variancia, a szórás legfontosabb összefüggéseiben

- n pozitív egész számot,
- $X,\ Y,\ Z,\ \ X_1, X_2, \ldots, X_n$ valószínűségi változókat,
- $a, b, c,\ a_1, a_2, \ldots, a_n$ valós számokat jelölnek.

Az elméleti összefügéseket valószínűségi változókra mondjuk ki. Remélhetőleg az Olvasó tisztában van vele, hogyan kell a valószínűségi változók egyes jellemzőit a valószínűségi változó eloszlásából meghatározni, és ezért a valószínűségi változókra kimondott összefüggéseket le tudja fordítani az eloszlások nyelvére is.

1. *Valószínűségi változó konstansszorosának a várható értéke:*

$$\mathrm{E}(c\,X) = c\,\mathrm{E}(X)$$

2. *A várható érték összegzési tulajdonsága:*
   - Két tagra:
   $$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$
   - Több tagra:
   $$\mathrm{E}(X_1 + X_2 + \ldots + X_n) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + \ldots + \mathrm{E}(X_n)$$

3. *A várható érték linearitási tulajdonsága:*
   - Két tagra:
   $$\mathrm{E}(a\ X + b\ Y) = a\ \mathrm{E}(X) + b\,\mathrm{E}(Y)$$
   - Több tagra:
   $$\mathrm{E}(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2) + \ldots + a_n\mathrm{E}(X_n)$$

4. *Összeg várható értéke, amikor a tagok várható értéke azonos:*
   Ha $X_1, X_2, \ldots, X_n$ valószínűségi változók mindegyikének a várható értéke μ, akkor
   $$\mathrm{E}(X_1 + X_2 + \ldots + X_n) = n\mu$$

5. *Átlag várható értéke, amikor a tagok várható értéke azonos:*
   Ha $X_1, X_2, \ldots, X_n$ valószínűségi változók mindegyikének a várható értéke μ, akkor
   $$\mathrm{E}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \mu$$

6. *Független valószínűségi változók szorzatának a várható értéke:*
   Ha X és Y független valószínűségi változók, akkor
   $$\mathrm{E}(\ X\ \cdot\ Y\ ) = \mathrm{E}(X)\ \cdot\ \mathrm{E}(Y)$$

Bizonyítás: Az 1-5. állítások kézenfekvőek. A 6. állítás bizonyítása: A $t(x, y) = x\,y$ függvényre alkalmazva a várható érték

$$\mathrm{E}\,(\,t(X,Y)\,) = \sum_{(x,y)} t(x,y)\;\;p(x,y)$$

általános képletét, ezt kapjuk:

$$\mathrm{E}\,(\,X\,Y\,) = \sum_{(x,y)} x\,y\;\;p(x,y) = \sum_{(x,y)} x\,y\;\;p_1(x)\,p_2(y) =$$

$$= \sum_{(x,y)} x\,p_1(x)\;\;y\,p_2(y) = \sum_{x} x\,p_1(x)\;\sum_{y} y\,p_2(y) = \mathrm{E}\,(\,X\,)\;\;\mathrm{E}\,(\,Y\,)$$

16.2. Variancia tulajdonságai

1. *Valószínűségi változó konstansszorosának a varianciája:*

$$\text{VAR}(cX) = c^2\;\text{VAR}(X)$$

2. *Független valószínűségi változók összegének a varianciája:*
   Ha X és Y független valószínűségi változók, akkor

$$\text{VAR}(X+Y) = \text{VAR}(X) + \text{VAR}(Y)$$

3. *Független valószínűségi változók összegének a varianciája, ha a tagok varianciája azonos:*
   Ha $X_1, X_2, \ldots, X_n$ független valószínűségi változók, melyek mindegyikének a varianciája σ^2, akkor

$$\text{VAR}(X_1 + X_2 + \ldots + X_n) = n\;\sigma^2$$

   Megjegyzés: Ha X és Y nem függetlenek, akkor ez a szabály általában nem igaz.

4. *Független valószínűségi változók átlagának a varianciája, ha a tagok varianciája azonos:*
   Ha $X_1, X_2, \ldots, X_n$ független valószínűségi változók, melyek mindegyikének a varianciája σ^2, akkor

$$\text{VAR}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

Bizonyítás: Csak a 2. állítás szorul magyarázatra, a többi kézenfekvő. A 2. állítás bizonyítása:

A variancia definícióját az $X+Y$ valószínűségi változóra alkalmazzuk:

$$\text{VAR}(X+Y) = \mathrm{E}\left(\left[\;X+Y\;-\;\mathrm{E}(\;X+Y\;)\;\right]^2\right)$$

Felhasználjuk, hogy $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$, és ezt kapjuk:

$$\text{VAR}(X+Y) = \mathrm{E}\left(\left[\;X+Y\;-\;\mathrm{E}(X)\;-\;\mathrm{E}(Y)\;\right]^2\right)$$

A szögletes zárójelen belül átrendezzük a tagokat:

$$\text{VAR}(X+Y) = \mathrm{E}\left(\left[\;(\;X-\mathrm{E}(X)\;)+(\;Y-\mathrm{E}(Y)\;)\;\right]^2\right)$$

Elvégezzük a négyzetre emelést:

$$\mathrm{VAR}(X+Y) = \mathrm{E}\Big((\ X - \mathrm{E}(X)\ )^2 + (\ Y - \mathrm{E}(Y)\ )^2 + \ 2\ \cdot\ (\ X - \mathrm{E}(X)\ )\ \cdot\ (\ Y - \mathrm{E}(Y)\ )\ \Big)$$

A három tag összegének a várható értékére három várható érték összegére esik szét:

$$\mathrm{VAR}(X+Y) = \mathrm{E}\Big((\ X - \mathrm{E}(X)\ )^2 \Big) + \mathrm{E}\Big((\ Y - \mathrm{E}(Y)\ )^2 \Big) + \ \mathrm{E}\,(\ 2\ \cdot\ (\ X - \mathrm{E}(X))\ \cdot\ (\ Y - \mathrm{E}(Y)\ )\ )$$

A harmadik tagban a függetlenség miatt a szorzat várható értéke egyenlő a várható értékek szorzatával,

$$\mathrm{VAR}(X+Y) = \mathrm{E}\Big((\ X - \mathrm{E}(X)\ )^2 \Big) + \mathrm{E}\Big((\ Y - \mathrm{E}(Y)\ )^2 \Big) + 2\ \cdot\ \mathrm{E}\,(\ X - \mathrm{E}(X)\ )\ \cdot\ \mathrm{E}\,(\ Y - \mathrm{E}(Y)\ )$$

Felhasználva, hogy

$$\mathrm{E}\,(\ X - \mathrm{E}(X)\ ) = 0 \ \text{ és } \ \mathrm{E}\,(\ Y - \mathrm{E}(Y)\ ) = 0$$

kiadódik, hogy a harmadik tag eltűnik:

$$\mathrm{VAR}(X+Y) = \mathrm{E}\Big((\ X - \mathrm{E}(X)\ )^2 \Big) + \mathrm{E}\Big((\ Y - \mathrm{E}(Y)\ )^2 \Big)$$

Itt pedig ráismerünk a varianciák összegére:

$$\mathrm{VAR}(X+Y) = \mathrm{VAR}(X) + \mathrm{VAR}(Y)$$

16.3. Szórás tulajdonságai

1. *Valószínűségi változó konstansszorosának a szórása:*

   $$\mathrm{SD}(c\,X) = |c|\ \mathrm{SD}(X)$$

2. *Független valószínűségi változók összegének a szórása:*
   Ha X és Y független valószínűségi változók, akkor

   $$\mathrm{SD}^2(X+Y) = \mathrm{SD}^2(X) + \mathrm{SD}^2(Y)$$

   azaz

   $$\mathrm{SD}(X+Y) = \sqrt{\mathrm{SD}^2(X) + \mathrm{SD}^2(Y)}$$

3. *Négyzetgyök szabály az összeg szórására:*
   Ha $X_1, X_2, \ldots, X_n$ független valószínűségi változók, melyek mindegyikének a szórása σ, akkor

   $$\mathrm{SD}(X_1 + X_2 + \ldots + X_n) = \sqrt{n}\ \sigma$$

4. *Négyzetgyök szabály az átlag szórására:*
   Ha $X_1, X_2, \ldots, X_n$ független valószínűségi változók, melyek mindegyikének a szórása σ, akkor

   $$\mathrm{SD}\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \right) = \frac{\sigma}{\sqrt{n}}$$

Egyik állítás sem szorul különösebb magyarázatra.

16.4. Gyakorló feladatok

1. Egy dobozban két cédula van, rajtuk az 1, illetve a 2 számok. Visszatevéssel húzogatunk, és egy papírra mindig felírjuk a kihúzott számot? Legyen X = ahány húzás kell ahhoz, hogy a papírunkon mindkét szám ott legyen. Határozza meg X

   (a) várható értékét!

   (b) szórását!

2. Egy dobozban három cédula van, rajtuk az 1, 2, illetve a 3 számok. Visszatevéssel húzogatunk, és egy papírra mindig felírjuk a kihúzott számot? Legyen X = ahány húzás kell ahhoz, hogy a papírunkon mind a három szám ott legyen. Határozza meg X

   (a) várható értékét!

   (b) szórását!

   (Segítség: Az első húzással kihúzunk valamit. Véletlentől függ, hogy ezután hány húzás kell ahhoz, hogy egy ettől különböző számot húzzunk. Jelölje a szükséges húzások számát Y. Véletlentől függ, hogy ezután hány húzás kell ahhoz, hogy a már kihúzott két számtól különböző számot húzzunk. Jelölje a szükséges húzások számát Z. Nem nehéz rájönni, hogy milyen eloszlást követ Y és Z, és hogy mennyi a várható értékük. Ha ezek után felhasználja a nyilvánvaló $X = 1 + Y + Z$ tényt, megkapja kérdezett várható értéket.)

3. Egy dobozban 10 cédula van 1 -től 10 -ig számozva. Visszatevéssel húzogatunk, és egy papírra mindig felírjuk a kihúzott számot? Legyen X = ahány húzás kell ahhoz, hogy a papírunkon az összes szám ott legyen. Határozza meg X

   (a) várható értékét!

   (b) szórását!

4. Tegyük fel, hogy egy eloszlás szerint

   (a) a 2 koordinátájú pontra eső valószínűség 0.3 . Az eloszlást csak annyiban változtatjuk meg, hogy a 2 ponton lévő 0.3 valószínűséget a 13 pontba helyezzük át. Mennyi lesz az új eloszlás várható értéke, ha az eredetié 5 volt?

   (b) az x_0 koordinátájú pontra eső valószínűség p_0. Az eloszlást csak annyiban változtatjuk meg, hogy az x_0 ponton lévő p_0 valószínűséget az $x_0 + b$ pontba helyezzük át. Mennyi lesz az új eloszlás várható értéke, ha az eredetié m volt? (b konstans)

   (c) az x_0 koordinátájú pontra eső valószínűség p_0. Az eloszlást csak annyiban változtatjuk meg, hogy az x_0 ponton lévő p_0 valószínűséget az $a\,x_0$ pontba helyezzük át. Mennyi lesz az új eloszlás várható értéke, ha az eredetié m volt? (a konstans)

   (d) az x_0 koordinátájú pontra eső valószínűség p_0. Az eloszlást csak annyiban változtatjuk meg, hogy az x_0 ponton lévő p_0 valószínűséget az $a\,x_0 + b$ pontba helyezzük át. Mennyi lesz az új eloszlás várható értéke, ha az eredetié m volt? (a és b konstansok)

17. Nevezetes eloszlások várható értéke, varianciája, szórása – formulák

17.1. Hipergeometrikus eloszlás

paraméterek: M, N, n

várható érték: $\mathrm{E}(X) = n\,\frac{M}{N}$

variancia: $VAR(X) = n\,\frac{N-n}{N-1}\,\frac{M}{N}\,(1-\frac{M}{N})$

szórás: $\mathrm{SD}(X) = \sqrt{n\,\frac{N-n}{N-1}\,\frac{M}{N}\,(1-\frac{M}{N})}$

módusz:

Ha $(n+1)\,\frac{M+1}{N+2}$ egész szám, akkor két módusz van: $(n+1)\,\frac{M+1}{N+2}$ és $(n+1)\,\frac{M+1}{N+2}-1$

Ha $(n+1)\,\frac{M+1}{N+2}$ nem egész szám, akkor egy módusz van: $\left\lfloor (n+1)\,\frac{M+1}{N+2} \right\rfloor$

17.2. Binomiális eloszlás

paraméterek: n és p

várható érték: $\mathrm{E}(X) = n\,p$

variancia: $VAR(X) = n\,p\,(1-p)$

szórás: $\mathrm{SD}(X) = \sqrt{n\,p\,(1-p)}$

módusz:

Ha $(n+1)\,p$ egész, akkor két módusz van: $(n+1)\,p$ és $(n+1)\,p-1$

Ha $(n+1)\,p$ nem egész, akkor egy módusz van: $\lfloor (n+1)\,p \rfloor$

17.3. Indikátor eloszlás

paraméter: p

várható érték: $\mathrm{E}(X) = p$

variancia: $VAR(X) = p\,(1-p)$

szórás: $\mathrm{SD}(X) = \sqrt{p\,(1-p)}$

módusz:

Ha $p = \frac{1}{2}$, akkor két módusz van a 0 és az 1

Ha $p < \frac{1}{2}$, akkor egy módusz van, a 0

Ha $p > \frac{1}{2}$, akkor egy módusz van, az 1

17.4. Optimista geometriai eloszlás

paraméter: p

várható érték: $\mathrm{E}(X) = \frac{1}{p}$

variancia: $VAR(X) = \frac{(1-p)}{p^2}$

szórás: $\mathrm{SD}(X) = \frac{\sqrt{(1-p)}}{p}$

módusz: 1

17.5. Pesszimista geometriai eloszlás

paraméter: p

várható érték: $\mathrm{E}(X) = \frac{1}{p} - 1 = \frac{1-p}{p}$

variancia: $\mathrm{VAR}(X) = \frac{(1-p)}{p^2}$

szórás: $\mathrm{SD}(X) = \frac{\sqrt{(1-p)}}{p}$

módusz: 0

17.6. Optimista negatív binomiális eloszlás

paraméterek: (r, p)

várható érték: $\mathrm{E}(X) = \frac{r}{p}$

variancia: $\mathrm{VAR}(X) = \frac{r\ (1-p)}{p^2}$

szórása: $\mathrm{SD}(X) = \frac{\sqrt{r\ (1-p)}}{p}$

módusza:

Ha $\frac{r-1}{p}$ egész, akkor két módusz van: $\frac{r-1}{p} + 1$ és $\frac{r-1}{p}$

Ha $\frac{r-1}{p}$ nem egész, akkor egy módusz van: $\left\lfloor \frac{r-1}{p} \right\rfloor + 1$

17.7. Pesszimista negatív binomiális eloszlás

paraméterek: (r, p)

várható érték: $\mathrm{E}(X) = \frac{r}{p} - r = r\,\frac{1-p}{p}$

variancia: $\mathrm{VAR}(X) = \frac{r\,(1-p)}{p^2}$

szórás: $\mathrm{SD}(X) = \frac{\sqrt{r\,(1-p)}}{p}$

módusz:

Ha $\frac{r-1}{p}$ egész, akkor két módusz van: $\frac{r-1}{p} - r + 1$ és $\frac{r-1}{p} - r$

Ha $\frac{r-1}{p}$ nem egész, akkor egy módusz van: $\left\lfloor \frac{r-1}{p} \right\rfloor - r + 1$

17.8. Poisson eloszlás

paraméter: λ

várható érték: λ

variancia: λ

szórás: $\sqrt{\lambda}$

módusz:

Ha λ egész, akkor két módusz van: λ és $\lambda - 1$

Ha λ nem egész, akkor egy módusz van: $\lfloor \lambda \rfloor$

Feladat: Telefonhívások. Tanszékünkre a délelőtti órákban óránként átlagosan 3.3 telefonhívás érkezik. Mi a valószínűsége annak, hogy

1. délelőtt 9 és 10 óra között pontosan 4 hívás érkezik?
2. délelőtt 9 és 11 óra között pontosan 4 hívás érkezik?
3. délelőtt 9 és 11 óra között 4 -nél több hívás érkezik?

Megoldás: Egy adott időintervallumban a hívások száma Poisson eloszlást követ. Egy óra alatt a várható érték 3.3, két óra alatt 6.6. Ezért

1.
$$\text{P(délelőtt 9 és 10 között pontosan 4 hívás)} =$$
$$= \frac{(3.3)^4}{4!}\,\mathrm{e}^{-3.3} = \texttt{POISSON(4 ; 3.3 ; FALSE)} = 0.18$$

2.

$$\text{P(délelőtt 9 és 11 között pontosan 4 hívás)} =$$
$$= \frac{(6.6)^4}{4!}\,\mathrm{e}^{-6.6} = \texttt{POISSON(4 ; 6.6 ; FALSE)} = 0.11$$

3.

$$\text{P(délelőtt 9 és 11 között 4 -nél több hívás)} =$$
$$= 1 - \text{P(délelőtt 9 és 11 között 4 vagy kevesebb hívás)} =$$
$$= 1 - \sum_{k=0}^{4} \frac{(6.6)^k}{k!}\,\mathrm{e}^{-6.6} = \texttt{1 - POISSON(4 ; 6.6 ; TRUE)} = 0.79$$

Feladat: Csillaghullás. Fogadjuk el, hogy augusztus közepén átlagoslan kb. 10 percenként lehet egy-egy csillaghullást látni. Mi a valószínűsége annak, hogy 15 perc alatt (pontosan) 2 csillaghullást lát valaki?

Megoldás: A 15 perc alatti csillaghullások száma Poisson eloszlást követ, hiszen sok a meteorit a világűrben, melyek egymással mit sem törődve száguldoznak, és minden meteoritra igaz, hogy a tekintett 15 perc alatt kis eséllyel lép be a fejünk felett a légkörbe. Mivel átlagosan 10 percenként jönnek a csillaghullások, 15 perc alatt a csillaghullások számának a várható értéke 1.5. A Poisson eloszlás paramétere 1.5. Ezért a kérdezett valószínűség:

$$\text{P(2 csillaghullás)} = \frac{(1.5)^2}{2!}\,\mathrm{e}^{-1.5} =$$
$$= \texttt{POISSON(2 ; 1.5 ; FALSE)} = 0.25$$

17.9. Gyakorló feladatok

1. Percenként átlagosan 2 hívás érkezik a tudakozóközpontba. Mi annak a valószínűsége, hogy 10:00 és 10:05 között legalább 4 hívás érkezik?

2. Átlagosan hány szem mazsolának kell lennie egy sütiben ahhoz, h ogy egy véletlenszerűen kiválasztott sütiben 99%-os valószínűséggel legyen (legalább egy szem) mazsola?

3. Egy dobozban 5 piros és 4 fehér golyó van. 3 golyót kiveszünk visszatevés nélkül. Legyen X a kihúzott piros golyók száma. Számolja ki X eloszlását, várható értékét, varianciáját, szórását!

4. Ha egy országban a átlagosan 2,7 hármasiker születik, akkor mi a valószínűsége annak, hogy pontosan 4 hármasiker születik

   (a) egy év alatt?

   (b) két év alatt?

5. Sok év statisztikája áll rendelkezésünkre arra nézve, hogy naponta hány lakástűz volt Budapesten. A napi négy tűzeset ugyanolyan relatív gyakorisággal fordul elő, mint az öt tűzeset. Becsülje meg, hogy a napok körülbelül hány százalékában fordul elő a két tűzeset!

6. Sok év adatai alapján feltesszük, hogy egy bizonyos kisvárosban naponta átlagosan 7.5 könnyű baleset és 2.8 súlyos baleset történik. Mennyi annak a valószínűsége, hogy egy nap alatt a könnyű balesetek száma

   (a) 10 -nél kisebb?

   (b) 5 -nél nagyobb?

   (c) 5 -nél nagyobb, de 10 -nél kisebb?

   (d) 10 -nél kisebb, feltéve, hogy 5 -nél nagyobb?

(e) 5 -nél nagyobb, feltéve, hogy 10 -nél kisebb?

Hogyan módosulnak a fenti kérdésekre adott válaszok, ha nem az egy, hanem a három egymást követő napon bekövetkező balesetek számára vonatkoznak?

18. Nevezetes eloszlások várható értékei – levezetések

18.1. Egyenletes eloszlás

$$\mathrm{E}(X) = \sum_x x\, p(x) = \sum_{x=A}^{B} x \; \frac{1}{B-A+1} =$$

$$\frac{1}{B-A+1} \; \sum_{x=A}^{B} x = \frac{1}{B-A+1} \; (B-A+1)\, \frac{A+B}{2} = \frac{A+B}{2}$$

Mivel az eloszláős szimmetrikus az $\frac{a+b}{2}$ pontra, nyilvánvaló, hogy a várható érték $\frac{a+b}{2}$.

18.2. Hipergeometrikus eloszlás *(Extra tananyag)*

$$\mathrm{E}(X) = \sum_x x\, p(x) =$$

$$\sum_{x=\max(0,n-B)}^{\min(n,A)} x \, \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} =$$

$$\sum_{x=\max(1,n-B)}^{\min(n,A)} x \, \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} =$$

$$\sum_{x=\max(1,n-B)}^{\min(n,A)} \frac{x\, \binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} =$$

$$\sum_{x=\max(1,n-B)}^{\min(n,A)} \frac{A\, \binom{A-1}{x-1}\binom{B}{n-x}}{\frac{A+B}{n}\binom{A-1+B}{n-1}} =$$

$$n\, \frac{A}{A+B} \; \sum_{x=\max(1,n-B)}^{\min(n,A)} \frac{\binom{A-1}{x-1}\binom{B}{n-x}}{\binom{A-1+B}{n-1}} =$$

$$n\, \frac{A}{A+B} \; \sum_{x=\max(0,n-1-B)}^{\min(n-1,A-1)} \frac{\binom{A-1}{y}\binom{B}{n-1-y}}{\binom{A-1+B}{n-1}} = n\, \frac{A}{A+B}$$

A levezetés során $x-1$ et y -nal helyettesítettük, és ennek megfelelően $1+y$ helyére x et irtunk. Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{x=\max(0,n-1-B)}^{\min(n-1,A-1)} \frac{\binom{A-1}{y}\binom{B}{n-1-y}}{\binom{A-1+B}{n-1}} = 1$$

ami következik abból a tényből, hogy az $A-1, B, n-1$ paraméterű hipergeometrikus eloszlás súlyfüggvénye:

$$\frac{\binom{A-1}{y}\binom{B}{n-1-y}}{\binom{A-1+B}{n-1}}$$

$$(\max(0, n-1-B) \le x \le \min(n-1, A-1))$$

18.3. Indikátor eloszlás

$$\mathrm{E}(X) = \sum_x x\, p(x) = 0(1-p) + 1p = p$$

18.4. Binomiális eloszlás

$$\mathrm{E}(X) = \sum_x x\, p(x) =$$

$$\sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=1}^{n} n \binom{n-1}{x-1}\, p\, p^{x-1}\, (1-p)^{n-x} =$$

$$np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} =$$

$$np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = np$$

A levezetés során $x-1$ et y -nal helyettesítettük, és ennek megfelelően $1+y$ helyére x et irtunk. Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = 1$$

ami következik abból a tényből, hogy az $n-1$ és p paraméterű binomiális eloszlás súlyfüggvénye:

$$\binom{n-1}{y} p^y (1-p)^{n-1-y} \quad \text{if} y = 0, 1, 2, \ldots, n$$

18.5. Geometriai eloszlás (optimista)

Két bizonyítást is adunk. Az első bizonyítás végtelen mértani sorok összegzésein alapul:

$$\mathrm{E}(X) = \sum_x x\, p(x) =$$

$$
\begin{array}{llllllllllll}
p & + & 2pq & + & 3pq^2 & + & 4pq^3 & + & \ldots & = \\
\\
p & + & pq & + & pq^2 & + & pq^3 & + & \ldots \\
 & + & pq & + & pq^2 & + & pq^3 & + & \ldots \\
 & & & + & pq^2 & + & pq^3 & + & \ldots \\
 & & & & & + & pq^3 & + & \ldots \\
 & & & & & & & \ddots \\
 & & & & & & & & & = \\
\frac{p}{1-q} \\
 & + & \frac{pq}{1-q} \\
 & & & + & \frac{pq^2}{1-q} \\
 & & & & & + & \frac{pq^3}{1-q} \\
 & & & & & & & \ddots \\
 & & & & & & & & & = \\
1 \\
 & + & q \\
 & & & + & q^2 \\
 & & & & & + & q^3 \\
 & & & & & & & \ddots \\
 & & & & & & & & & =
\end{array}
$$

$$\frac{1}{1-q} = \frac{1}{p}$$

A második bizonyítás a hatvány sorok elméletén alapszik. A $q = 1 - p$ jelöléssel élve ezt kapjuk:

$$\mathrm{E}(X) = \sum_x \, x \, p(x) =$$

$$1 \; p + 2 \; p(1-p) + 3 \; p(1-p)^2 + 4 \; p(1-p)^3 + \ldots =$$

$$1 \; p + 2 \; pq + 3 \; pq^2 + 4 \; pq^3 + \ldots =$$

$$p \; \left(1 + 2q + 3q^2 + 4q^3 + \ldots\right) = p \; \frac{1}{(1-q)^2} = p \; \frac{1}{p^2} = \frac{1}{p}$$

Felhasználtuk a

$$1 + 2q + 3q^2 + 4q^3 + \ldots = \frac{1}{(1-q)^2}$$

azonosságot, amit az alábbiakban le is vezetünk úgy, hogy először felismerjük, hogy az adott sor egy mértani sor deriváltjaként fogható fel, aztán vesszük a végtelen mértani sor összegképletét, és végül azt deriváljuk:

$$1 + 2q + 3q^2 + 4q^3 + \ldots =$$

$$\frac{d}{dq} \left(1 + q + q^2 + q^3 + q^4 \ldots\right) =$$

$$\frac{d}{dq} \left(\frac{1}{1-q}\right) = \frac{d}{dq} \left((1-q)^{-1}\right) = (1-q)^{-2} = \frac{1}{(1-q)^2}$$

18.6. Geometriai eloszlás (pesszimista)

Mivel a pesszimista geometriai eloszlás az optimistából úgy adódik, hogy azt 1 egységgel balra toljuk, a pesszimista geometriai eloszlás várható értéke nem más, mint az optimista geometriai eloszlás várható értéke mínusz 1.

$$\mathrm{E}(X) = \frac{1}{p} - 1$$

18.7. Negatív binomiális eloszlás (optimista) *(Extra tananyag)*

$$\mathrm{E}(X) = \sum_x \; x \, p(x) =$$

$$\sum_{x=r}^{\infty} x \binom{x-1}{r-1} p^r (1-p)^{x-r} =$$

$$\sum_{x=r}^{\infty} r \binom{x}{r} \frac{p^{r+1}}{p} (1-p)^{x-r} =$$

$$\frac{r}{p} \sum_{x=r}^{\infty} \binom{x}{r} p^{r+1} (1-p)^{x-r} =$$

$$\frac{r}{p} \sum_{x=r+1}^{\infty} \binom{y-1}{r} p^{r+1} (1-p)^{y-1-r} =$$

$$\frac{r}{p}$$

$$\sum_{x=r+1}^{\infty} \binom{y-1}{r} p^{1+r} (1-p)^{y-1-r} = 1$$

ami következik abból a tényből, hogy a $r+1$ és p paraméterű optimista negatív binomiális eloszlás súlyfüggvénye

$$\binom{y-1}{r} p^{1+r} (1-p)^{y-1-r} \quad \textbf{if } y = r+1, r+2, r+3, \ldots$$

18.8. Negatív binomiális eloszlás (pesszimista) *(Extra tananyag)*

Mivel a pesszimista negatív binomiális eloszlás az optimistából úgy adódik, hogy azt r egységgel balra toljuk, a pesszimista negatív binomiális eloszlás várható értéke nem más, mint az optiimista geometriai eloszlás várható értéke mínusz r

$$\mathrm{E}(X) = \frac{r}{p} - r$$

18.9. Poisson eloszlás

$$\mathrm{E}(X) = \sum_x \; x \, p(x) =$$

$$\sum_{x=0}^{\infty} x \;\; \frac{\lambda^x}{x!} e^{-\lambda} =$$

$$\sum_{x=1}^{\infty} x \;\; \frac{\lambda^x}{x!} e^{-\lambda} =$$

$$\sum_{x=1}^{\infty} \lambda \; \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} =$$

$$\lambda \;\; \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} =$$

$$\lambda \;\; \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} = \lambda$$

A levezetés során $x-1$ et y -nal helyettesítettük, és ennek megfelelően $1+y$ helyére x et irtunk. Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} = 1$$

ami következik abból a tényból, hogy a λ paraméterű Poisson eloszlás súlyfüggénye

$$\frac{\lambda^y}{y!} e^{-\lambda} \quad (y = 0, 1, 2, \ldots)$$

19. Binomiáls eloszlás második momentumának, varianciájának és szórásának levezetése

19.1. Második momentum

Állítás:

$$\mathrm{E}(X^2) = np \; + \; n^2p^2 \; - \; np^2$$

Bizonyítás *(Extra tananyag)*:

$$\mathrm{E}(X^2) = \sum_x \; x^2 \; p(x) =$$

$$\sum_{x=0}^{n} \; x^2 \; \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=1}^{n} \; x^2 \; \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=1}^{n} \; n \; x \; \binom{n-1}{x-1} \; p^x \; (1-p)^{n-x} =$$

$$np \; \sum_{x=1}^{n} \; x \; \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} =$$

Most $(x-1)$ -et y -nal helyettesítjük, és így $(1+y)$ -t írunk x helyére:

$$np \; \sum_{y=0}^{n-1} (1+y) \; \; \binom{n-1}{y} p^y (1-p)^{n-1-y} =$$

Az összeg két összegre bontható:

$$np \; \left[\left(\sum_{y=0}^{n-1} 1 \; \; \binom{n-1}{y} p^y (1-p)^{n-1-y} \right) + \left(\sum_{y=0}^{n-1} y \; \; \binom{n-1}{y} p^y (1-p)^{n-1-y} \right) \right] =$$

Itt a szögletes zárójal belsejében álló első szumma az $n-1$ és p paraméterű binomiláis eloszlás tagjainak az összege, ami 1. A második szumma az $n-1$ és p paraméterű binomiláis eloszlás várható értéke, ami $(n-1)p$. Ezt kapjuk:

$$np \;\; [\, 1 \; + \; (n-1) \; p \,] = np \; + \; n^2p^2 \; - \; np^2$$

19.2. Variancia és szórás

Állítás:

$$\mathrm{VAR}(X) = np(1-p)$$
$$\mathrm{SD}(X) = \sqrt{np(1-p)}$$

Bizonyítás. *(Extra tananyag.)* Mint már tudjuk, a binomiális eloszlás második momentuma

$$\mathrm{E}(X^2) = n^2p^2 - np^2 + np$$

várható értéke

$$\mathrm{E}(X) = np$$

Ezért a varianciája

$$\mathrm{VAR}(X) = \left(n^2p^2 - np^2 + np\right) - (np)^2 = np - np^2 = np(1-p)(1-p)$$

szórása pedig

$$\mathrm{SD}(X) = \sqrt{np(1-p)}$$

20. Feltételes várható érték, variancia, szórás

Először a feltételes várható érték fogalmával ismerkedünk meg, azután jön a többi feltételes "izé".

20.1. Feltételes várható érték

Ha X és Y diszkrét valószínűségi változók kapcsán korábban bevezettük a $p_{2|1}(y|x)$, $p_{1|2}(x|y)$ feltételes súlyfüggvényeket (feltételes eloszlásokat). Most a feltételes súlyfüggvényekre támaszkodva bevezettjük először a feltételes várható érték fogalmát.

A feltételes súlyfüggvényből számított

$$\sum_{y=-\infty}^{\infty} y \, p_{2|1}(y|x)$$

várható értéket **feltételes várható érték**nek nevezzük. Az $X = x$ feltétel mellett az Y feltételes várható értékét $\mathrm{E}(Y|X = x)$ -szel jelöljük. Tehát

$$\mathrm{E}(Y|X = x) = \sum_{y=-\infty}^{\infty} y \, p_{2|1}(y|x)$$

Az $X = x$ feltétel mellett az Y feltételes várható értéke azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y -ra kapott kísérleti eredmények átlaga.

Hasonlóképpen az $Y = y$ feltétel mellett az X feltételes várható értékét $\mathrm{E}(X|Y = y)$ -nal jelöljük. Nyilván

$$\mathrm{E}(X|Y = y) = \sum_{x=-\infty}^{\infty} x \, p_{1|2}(x|y)$$

Az $Y = y$ feltétel mellett az X feltételes várható értéke azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $Y = y$, akkor az érvényes kísérletekben körülbelül mennyi az X értékek átlaga.

20.2. Feltételes variancia

A feltételes súlyfüggvényből számított varianciát **feltételes varianciá**nak nevezzük.

Az $X = x$ feltétel mellett az Y feltételes varianciája azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y -ra kapott kísérleti eredmények varianciája.

Az $Y = y$ feltétel mellett az X feltételes varianciája azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $Y = y$, akkor az érvényes kísérletekben körülbelül mennyi az X -re kapott kísérleti eredmények varianciája.

20.3. Feltételes szórás

A feltételes súlyfüggvényből számított szórást **feltételes szórás**nak nevezzük.

Az $X = x$ feltétel mellett az Y feltételes szórása azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y -ra kapott kísérleti eredmények szórása.

Az $Y = y$ feltétel mellett az X feltételes varianciája aztt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $Y = y$, akkor az érvényes kísérletekben körülbelül mennyi az X -re kapott kísérleti eredmények szórása.

Megjegyzés: A feltételes súlyfüggvényből számított "izé"-t **feltételes "izé"** -nek nevezzük. Az $X = x$ feltétel mellett az Y feltételes "izé" -je azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y -ra kapott kísérleti eredmények "izé" -je.

21. Vegyes feladatok diszkrét eloszlásokra

1. Egy 30 fős osztályban 17 lány van. Véletlenszerűen kiválasztanak az osztályból egy 12 fős csapatot egy vetélkedőre. Legyen a csapatba került lányok száma X. Kérdés: $\mathrm{P}(X = 7) =?$
2. Egy 400 oldalas könyvben összesen 200 sajtóh**u**ba van (véletlenszerűen elszórva). Mennyi a valószínűsége annak, hogy a 13. oldalon több, mint egy sajtóh**u**ba van? Hány sajtóh**u**ba a legvalószínűbb a 13. oldalon? Mennyi a valószínűsége annak, hogy a 13. és a 14. oldalon együtt több, mint két sajtóh**u**ba van?
3. 80 üveg bor van egy borospincében össze-vissza lerakva, ebből 30 fehér, 50 vörös. A vendégek a fogadóstól 3 üveg fehér és 7 vörösbort rendelnek, de a pincében kiégett a villany. A fogadós véletlenszerűen kiválaszt 10 üveget. Mi a valószínűsége, hogy minden vendég kap neki megfelelő itókát?
4. A vidámparkban a céllövöldében játszom. Egymás után vonulnak fel a célpontok, mindegyiket egymástól függetlenül 2/3 valószínűséggel eltalálom. Mi a valószínűsége, hogy 6 célzásból
   (a) pontosan 4 -et találok el?
   (b) 2 -nél többet találok el, de azért nem az összeset?
5. Blicc úr minden nap villamossal megy dolgozni, de nincs bérlete, sem jegye. A villamosra minden nap 0,2 valószínűséggel száll fel ellenőr, és ilyenkor 0,95 valószínűséggel elkapja Blicc urat. (Az ellenőr minden nap az addigiaktól függetlenül dönti el, ellenőrzi-e aznap Blicc úr villamosát.)
   (a) Mennyi a valószínűsége, hogy Blicc úrnak "szerencsés hete" van, azaz az 5 munkanap egyikén sem kell büntetést fizetnie?
   (b) Mennyi a valószínűsége, hogy pontosan kétszer kapják el egy hét munkanapjai alatt?
   (c) Feltéve, hogy Blicc úrnak "szerencsés hete" volt, mi a valószínűsége, hogy mind az ötször volt ellenőr a villamoson?
   (d) Mi a valószínűsége hogy csütörtökön büntetik meg másodszor?
6. Egy szöcske elindul a számegyenes origójából. Minden lépésnél 1/2 valószínűséggel jobbra, 1/2 valószínűséggel balra ugrik. 20 ugrás megtétele után
   (a) milyen valószínűséggel lesz a 0-ban?
   (b) milyen valószínűséggel lesz az 1-ben?
   (c) milyen valószínűséggel lesz a (-2)-ben, ha az utolsó előtti ugrás után a (-3)-ban volt?
7. Van két érmém, az egyik igazságos érme, a másik cinkelt, de ránézésre nem tudom őket megkülönböztetni egymástól. A cinkelt érme 3/4 valószínűséggel mutat fejet. Előveszem az egyik érmét a zsebemből, 1/2 eséllyel az igazságosat, 1/2 eséllyel a cinkeltet. A kiválasztott érmét feldobom 30-szor, és azt tapasztalom, hogy 25-ször mutatott fejet. Mi a valószínűsége, hogy a cinkelt érmét vettem elő?
8. Addig dobunk két kockával, amíg a két kockán lévő számjegyek összege 12 nem lesz.
   (a) Mennyi annak a valószínűsége, hogy pontosan nyolcszor dobunk 12-nél kisebb összeget, mielőtt 12-t dobnánk?
   (b) Mennyi a valószínűsége, hogy összesen nyolcszor dobunk?
9. Egy (szabálytalan) pénzérmét dobunk fel annyiszor, amíg fejet nem kapunk. Ha a fej dobás valószínűsége p, akkor mennyi a valószínűsége, hogy
   (a) pont k -szor dobunk a fej előtt?
   (b) pont k -szor dobunk az érmével?

10. Dobogatok a kockával és vonásal számolom, hogy hány hatost dobtam. Mi a valószínűsége, hogy a 12. dobásra húzom a harmadik vonást? Ha azt számolnám ki, hogy mennyi a valószínűsége, hogy 12-szer dobok hatostól különbözőt, mire kidobom a harmadik hatost, akkor különbözne ez az előző eredménytől ?

11. 100 kulcs közül csak 1 nyitja az előttünk lévő ajtót. A sötétben nem látjuk, hogy melyik kulcsot próbáltuk már ki, így a próbálgatások során többször is a kezünkbe kerülhet ugyanaz kulcs. Mi a valószínűsége, hogy legfeljebb 50 próbálkozással kinyitjuk az ajtót? És ha a kipróbált kulcsokat félretesszük?

12. Egymás után kérdezgetjük az embereket a születésnapjukról: melyik hónap hányadikán születtek.

    (a) Hányadik embernél adódik az első olyan születésnap, ami már korábban szerepelt? Határozza meg ennek a valószínűségi változónak az eloszlását: adja meg a súlyfüggvénynek a képletét!

    (b) Hányadik embernél adódik a második olyan születésnap, ami már korábban szerepelt? Határozza meg ennek a valószínűségi változónak az eloszlását: adja meg a súlyfüggvénynek a képletét!

13. Adja meg az alábbi valószínűségi változók eloszlását! Két szabályos dobókockával dobunk, és megfigyeljük a dobott számok

    (a) összegét.

    (b) különbségét.

14. Valaki minden héten egyetlen ötös lottó szelvénnyel játszik. Legalább hány hétig kell játszania ahhoz, hogy a hármas, négyes, ötös valószínűsége legalább 1/2 legyen? (Ez 3 különálló kérdés.)

15. Van két érmém, az egyik igazságos érme, a másik cinkelt, de ránézésre nem tudom őket megkülönböztetni egymástól. A cinkelt érme 3/4 valószínűséggel mutat fejet. Előveszem az egyik érmét a zsebemből, 1/2 eséllyel az igazságosat, 1/2 eséllyel a cinkeltet, és odaadom a hallgatóknak. 30 dobás után el kell dönteniük, melyik érme volt, amit elővettem. Hol húznák meg a döntési határt? (A 30 dobás közül hány fej esetén tippelnének még az igazságos érmére, hány fej esetén tippelnének már a cinkeltre?)

16. Egy dobozban N darab cédula van 1-től N -ig megszámozva. Visszatevés nélkül húzunk n -szer, majd a kihúzott számokat nagyság szerint sorba rakjuk. Tekintsük a nagyság szerinti

    (a) legkisebbet,

    (b) legnagyobbat,

    (c) 2-ik legkisebbet,

    (d) 3-ik legkisebbet,

    (e) s -edik legkisebbet.

    Határozza meg ezeknek a valószínűségi változóknak az eloszlását: adja meg a súlyfüggvény képletét!

17. *Általánosítás:* Egy dobozban A darab piros és B darab fehér golyó van. Visszatevés nélkül húzok az r -ik pirosig. Adjuk meg a súlyfüggvény és az eloszlásfüggvény képletét! Adja meg a súlyfüggvényt visszatevéses húzás esetén is!